Behavior Research Methods
2009, 41 (1), 61-74
doi:10.3758/BRM.41.1.61

Detection of interactions between a dichotomous moderator and a continuous predictor in moderated multiple regression with heterogeneous error variance

Gwowen Shieh
National Chiao Tung University, Hsinchu, Taiwan

Moderated multiple regression (MMR) has been widely used to investigate the interaction or moderating effects of a categorical moderator across a variety of subdisciplines in the behavioral and social sciences. In view of the frequent violation of the homogeneity of error variance assumption in MMR applications, the weighted least squares (WLS) approach has been proposed as one of the alternatives to the ordinary least squares method for the detection of the interaction effect between a dichotomous moderator and a continuous predictor. Although the existing result is informative in assuring the statistical accuracy and computational ease of the WLS-based method, no explicit algebraic formulation and underlying distributional details are available. This article aims to delineate the fundamental properties of the WLS test in connection with the well-known Welch procedure for regression slope homogeneity under error variance heterogeneity. With elaborately systematic derivation and analytic assessment, it is shown that the notion of WLS is implicitly embedded in the Welch approach. More importantly, extensive simulation study is conducted to demonstrate the conditions in which the Welch test will substantially outperform the WLS method; they may yield different conclusions. Welch's solution to the Behrens–Fisher problem is so entrenched that the use of its direct extension within the linear regression framework can arguably be recommended. In order to facilitate the application of Welch's procedure, the SAS and R computing algorithms are presented. The study contributes to the understanding of methodological variants for detecting the effect of a dichotomous moderator in the context of moderated multiple regression. Supplemental materials for this article may be downloaded from brm.psychonomic-journals.org/content/supplemental.

Researchers are often interested in determining whether the direction or strength of the relation between a predictor variable and a response variable varies with the value of a third or moderator variable. The existence of moderating effects implies that the predictor has a fundamentally distinct impact on the response across levels of the moderator. The formulation of differential prediction behavior occurs in diverse research settings such as gender studies. Essentially, the moderated relationships can be conceptualized and analyzed in terms of interaction effects between the predictor and moderator variables. It is consensually recognized that moderated multiple regression (MMR) has become the major technique for testing hypotheses about moderating effects of categorical variables in psychology, management, education, and related disciplines; see Aguinis (2004) for general and illuminating expositions. When the null hypothesis of no moderating effect is rejected, it indicates that the predictor–response relationship is stronger for one moderator-based group than for another. Neglect of an interaction effect or failure to detect a moderating effect generally leads to prediction bias in favor of subjects in some groups and against members of the other groups.

The procedure for detecting the effects of categorical moderator variables is methodologically identical to that for testing the equality of regression slope coefficients in two or more regression lines. Accordingly, the test can be conducted with the ordinary least square (OLS) partial F test in traditional MMR analysis. However, numerous studies have shown that MMR may often yield erroneous conclusions; in particular, many theory-based hypotheses of moderated phenomena are frequently not supported. In response to the failures to detect sound hypothesized moderating effects, several researchers have investigated the accuracy of MMR to evaluate moderating effects under various conditions. For example, Aguinis and Stone-Romero (1997) and Stone-Romero, Alliger, and Aguinis (1994) provided thorough treatments of the methodological artifacts and statistical implications associated with the effects of dichotomous moderators. More specifically, considerable attention has been devoted to raising awareness of the often violated assumption of homogeneous error variance when assessing moderating effects of categorical variables; see Aguinis, Petersen, and Pierce (1999) and Aguinis and Pierce (1998) for comprehensive descriptions and excellent reviews. It should be noted that

G. Shieh, gwshieh@mail.nctu.edu.tw

the 12-year review of Aguinis et al. showed that the violation of homogeneity assumption is approximately 40% to 60% of the MMR tests reported in three prestigious journals with rigorous methodological standards: *Academy of Management Journal*, *Journal of Applied Psychology*, and *Personnel Psychology*. Hence, they suggested that the violation situation is at least as common for tests reported in other journals in organizational science.

Naturally, the accuracy of the OLS-based F test depends on the strong assumption of homogeneous within-group error variance. Several empirical studies have been conducted to ascertain the effects of heterogeneous error variance on the performance of the F test. For detailed discussions, see Alexander and DeShon (1994), DeShon and Alexander (1994, 1996), Dretzke, Levin, and Serlin (1982), and Overton (2001). These Monte Carlo simulations concluded that the Type I error rate and power of the regular MMR F test may be substantially affected when group sample sizes are equal, and severely distorted when group sample sizes are unequal. Thus, researchers may commit a Type I error or a Type II error, depending on the specific sample characteristics and postulated model formulations. Consequently, the study may discover a fake interaction effect (Type I error) or mistakenly dismiss an important moderator variable (Type II error). In either case, the result impedes theoretical development and scientific advancement of moderation research. Apparently, the regular F test is not a proper procedure that accounts for the nature of heterogeneity of within-group error variance, and continual efforts have explored alternative methods when testing hypotheses about categorical moderators. The complexity of heterogeneity error variance incurs numerous investigations, which offer various approximations and computing algorithms for solving the problem.

Since the problem of heterogeneity of error variance in testing for equality of regression slopes is statistically equivalent to the problem of variance heterogeneity in ANOVA, the available tests for examining mean equality under variance heterogeneity in ANOVA can be applied to the detection of categorical moderating effects in MMR. Notably, the methods of Alexander and Govern (1994), James (1951), and Welch (1951) for testing the equality of K (≥ 2) independent means under heterogeneity of variance have been adapted to the tests for the equality of K independent regression slopes under heterogeneous error variance in Alexander and DeShon (1994), DeShon and Alexander (1994, 1996), and Dretzke et al. (1982). The exact formulations and test procedures of the three approximations A, J, and F^* developed from Alexander and Govern (1994), James (1951), and Welch (1951), respectively, can be found in DeShon and Alexander (1996). On the basis of the numerical examinations for two-group situation in the abovementioned studies, the performance of the three methods A, J, and F^* was essentially equivalent for $K = 2$, and choosing the best approximation is difficult. Nonetheless, the A approach was shown to possess a number of desirable characteristics and was recommended by DeShon and Alexander (1996) as the general procedure in lieu of regular MMR F statistics for testing categorical moderator hypotheses when the assumption on homogeneity of within-group error variance is not tenable. Although programs for computing A, J, and F^* are available in Aguinis et al. (1999) and DeShon and Alexander (1996), it was pointed out in Overton (2001) that these test procedures do not support follow-up analyses. Accordingly, Overton proposed a weighted least squares (WLS) approach for the $K = 2$ groups that maintains MMR within the familiar multiple regression framework. Moreover, it was demonstrated in Overton that the WLS F_ω test is not only accurate, but can also be easily executed using the standard procedures of the SAS statistical package. Therefore, one advantage of the WLS-based method over the existing approaches is that the corresponding follow-up analysis of moderating effects can be readily performed using the embedded features of SAS or other popular software systems. Consequently, there appears to be a lack of consensus in the literature on which method is most appropriate for detecting the effects of a dichotomous moderator variable in MMR under heterogeneous error variance.

Although the notion of the WLS procedure and its corresponding computing aspect are thoroughly presented in Overton (2001), no explicit analytical form of the test statistics was available. Even though Overton noted that the WLS-based MMR and the OLS-based MMR yield identical regression coefficient estimators but differ in their standard errors for the coefficient estimators, no further detailed expressions were given. It is of both methodological importance and practical interest to obtain the exact formulations of the coefficient estimators, associated estimated variance, and resulting test statistic in the context of the WLS principle. To our knowledge, no research to date has examined the theoretical issue of WLS-based procedures in greater detail. On the other hand, the hypothesis testing procedure of F-type statistics for assessing the moderating effects of a dichotomous variable is nondirectional in nature. Depending on the purpose of the study, a particular one-sided test might be preferable. Hence, it is more flexible and informative to conduct the test with a t statistic, since it can be used for one-sided alternatives, whereas a partial F test cannot. Furthermore, a confidence interval may be more useful for interpreting the magnitude of the moderating effect. As in many applications, a t statistic can be naturally adopted to construct a confidence interval. However, this is apparently not the case for an F test.

Since the detection of slope differences between two regression lines or interactions between a dichotomous moderator and a continuous predictor represents the vast majority of MMR research, this article attempts to derive the analytical results for the WLS analysis and the related criteria in a unified and relatively transparent way. The general formulations of the aforementioned techniques for comparing the equality of regression slopes of two and more groups are appealing and yet may not be necessary and advantageous for the current focus on the problem of the two-group situation. As discussed earlier, the nonsquared form of a partial F statistic or a partial t expression proves to be more versatile than the squared form of the F test in this particular MMR application. The examination of the established results helps strengthen the importance of the problem and reveals the closer functional relation between the existing vital approaches. It will be explicitly shown later that the WLS-based methods are closely related to the well-known Welch procedures. In the process, we also hope to account for some important findings and ambiguous issues that may have been overlooked in the literature.

The rest of the article is organized as follows. The next section describes the fundamental theory and analytical results for the inference of interactions between a dichotomous moderator and a continuous predictor in the context of MMR with heterogeneous error variance. Then the emphasis is placed on the underlying similarities and differences between the Welch and WLS methods. Extensive numerical investigations are conducted to exemplify the critical and subtle discrepancy between the two approaches. In order to enhance the application of the prevalent Welch procedure, the SAS and R programs are provided to ease the inferential analyses of hypothesis testing imposed by the technique.

Relationship Between Welch and WLS Procedures

Consider the following two simple linear regression models of the form

$$Y_{1i} = \beta_{10} + X_{1i}\beta_{11} + \varepsilon_{1i} \text{ and } Y_{2j} = \beta_{20} + X_{2j}\beta_{21} + \varepsilon_{2j}, \quad (1)$$

where ε_{1i} and ε_{2j} are *iid* $N(0, \sigma_1^2)$ and $N(0, \sigma_2^2)$ random variables, respectively, $i = 1, \ldots, N_1$, and $j = 1, \ldots, N_2$. In view of moderated multiple regression with the focus on the two-group parallelism problem, it is often more illuminating to combine the two models in Equation 1 as the following multiple regression model with a dichotomous moderator variable:

$$Y_k = \beta_{20} + X_k\beta_{21} + Z_k\delta_0 + Z_kX_k\delta_1 + \varepsilon_k,$$
$$k = 1, \ldots, N, N = N_1 + N_2, \quad (2)$$

where

$$\delta_0 = \beta_{10} - \beta_{20}, \delta_1 = \beta_{11} - \beta_{21};$$

$$Y_k = Y_{2j}, X_k = X_{2j}, \varepsilon_k = \varepsilon_{2j}, \text{ and } Z_k = 0 \text{ if } k = j, j = 1, \ldots, N_2;$$

$$Y_k = Y_{1i}, X_k = X_{1i}, \varepsilon_k = \varepsilon_{1i}, \text{ and } Z_k = 1 \text{ if } k = N_2 + i, i = 1, \ldots, N_1.$$

Methodologically, the existence and magnitude of regression coefficient $\delta_1 = \beta_{11} - \beta_{21}$, representing the influence of the moderating effect, is the major concern for the analysis of the moderated multiple regression model in Equation 2. In the present section, we discuss the statistical tests for the homogeneity of slopes of two simple regression models by summarizing the fundamental results from different disciplines; this not only underscores the importance of the problem, but also provides a comprehensive review of various solutions for moderation analysis. Also, more importantly, detailed analytical examinations are conducted to demonstrate the relationship and discrepancy of existing prominent methods. Conceivably, the well-supported recommendations offered in the presentation may be useful for empirical research.

Subgroup OLS

The fundamental statistical results are well known for the two simple regression models with normal error assumption given in Equation 1—for example, see Rencher (2000). Suppose that $\hat{\beta}_{10}, \hat{\beta}_{11}, \hat{\beta}_{20}$, and $\hat{\beta}_{21}$ are the subgroup OLS estimators of β_{10}, β_{11}, β_{20}, and β_{21}, respectively. Although it is not necessary to apply matrix algebra to derive these estimators for the two simple linear models, the matrix formulations presented in Appendix A will later be shown to be useful for demonstrating the relationship between WLS and OLS estimators.

We are especially concerned with the statistical properties associated with the two estimators of slope coefficients $\hat{\beta}_{11}$ and $\hat{\beta}_{21}$, specifically,

$$\hat{\beta}_{11} \sim N(\beta_{11}, \sigma_1^2/SSX_1) \text{ and } \hat{\beta}_{21} \sim N(\beta_{21}, \sigma_2^2/SSX_2),$$

where $SSX_1 = \Sigma_{i=1}^{N_1}(X_{1i} - \bar{X}_1)^2$ and $SSX_2 = \Sigma_{j=1}^{N_2}(X_{2j} - \bar{X}_2)^2$; $\bar{X}_1$ and $\bar{X}_2$ are the respective sample means of the X_{1i} and X_{2j} observations. For inferential purposes, $\hat{\sigma}_1^2 = SSE_1/(N_1 - 2)$ and $\hat{\sigma}_2^2 = SSE_2/(N_2 - 2)$ are the usual unbiased estimators of σ_1^2 and σ_2^2. Note that $\hat{\beta}_{11} - \hat{\beta}_{21} \sim N(\beta_{11} - \beta_{21}, \sigma_1^2/SSX_1 + \sigma_2^2/SSX_2)$. Moreover, the error sum of squares $SSE_1 \sim \sigma_1^2\chi^2(N_1 - 2)$ and $SSE_2 \sim \sigma_2^2\chi^2(N_2 - 2)$, where $\chi^2(N_1 - 2)$ and $\chi^2(N_2 - 2)$ are chi-square distributions with $N_1 - 2$ and $N_2 - 2$ degrees of freedom, respectively.

Welch Procedures

As a straightforward extension of Welch's well-known "approximate degrees of freedom" or "approximate t" solution to the Behrens–Fisher problem of comparing the difference between two means, Welch (1938, p. 356) also described the methods for comparing the difference in two regression slope coefficients within the simple regression framework of Equation 1. For the purpose of testing the hypothesis H_0: $\beta_{11} = \beta_{21}$, one of the methods proposed by Welch (1938) is to consider the approximate distribution $V \dot{\sim} t(\nu)$ when $\beta_{11} = \beta_{21}$, where

$$V = \frac{\hat{\beta}_{11} - \hat{\beta}_{21}}{\left(\hat{\sigma}_1^2 / SSX_1 + \hat{\sigma}_2^2 / SSX_2\right)^{1/2}}, \quad (3)$$

and $t(\nu)$ is a t distribution with degrees of freedom $\nu = \nu(\sigma_1^2/SSX_1, \sigma_2^2/SSX_2)$ with

$$1/\nu = \frac{1}{N_1 - 2}\left\{\frac{\sigma_1^2 / SSX_1}{\sigma_1^2 / SSX_1 + \sigma_2^2 / SSX_2}\right\}^2 + \frac{1}{N_2 - 2}\left\{\frac{\sigma_2^2 / SSX_2}{\sigma_1^2 / SSX_1 + \sigma_2^2 / SSX_2}\right\}^2.$$

Consequently, the unknown parameters σ_1^2 and σ_2^2 in ν are replaced by $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ for the practical purpose of hypothesis testing. Hence, it leads to the modified approximation with random number of degrees of freedom:

$$V \stackrel{\cdot}{\sim} t(\hat{\nu}), \tag{4}$$

where

$$1/\hat{\nu} = \frac{1}{N_1 - 2}\left\{\frac{\hat{\sigma}_1^2 / SSX_1}{\hat{\sigma}_1^2 / SSX_1 + \hat{\sigma}_2^2 / SSX_2}\right\}^2 + \frac{1}{N_2 - 2}\left\{\frac{\hat{\sigma}_2^2 / SSX_2}{\hat{\sigma}_1^2 / SSX_1 + \hat{\sigma}_2^2 / SSX_2}\right\}^2.$$

The null hypothesis is rejected at the significance level α if

$$|V| > t_{\hat{\nu},\alpha/2}, \tag{5}$$

where $t_{\hat{\nu},\alpha/2}$ is the $100(1 - \alpha/2)$ percentile of the t distribution $t(\hat{\nu})$ with degrees of freedom $\hat{\nu}$.

Several studies have shown that Welch's approximate degrees of freedom approach offers a reasonably accurate solution to the Behrens–Fisher problem; for example, see Best and Rayner (1987), Davenport and Webster (1975), Nel, van der Merwe, and Moser (1990), Scariano and Davenport (1986), and Scheffé (1970). Because the prescribed Welch's V for comparing regression slope coefficients is a natural adaptation of Welch's original approximate t test for equality of two normal means, the V test (Equation 5) should possess the same advantage of accurate control of the magnitudes of size and power. Nonetheless, it can be demonstrated that the general Welch–Aspin F^* test for comparing regression slope equality presented in DeShon and Alexander (1996) reduces to the F^* method in Dretzke et al. (1982) under the two-group circumstance. Moreover, the F^* of Dretzke et al. is actually the square of the V given in Equation 5, $F^* = V^2$. Consequently, F^* is referred to the F distribution with degrees of freedom 1 and $\hat{\nu}$, according to the notations used here. Notably, the accurate performance of the Welch procedure for comparing the slopes of two regression lines has been demonstrated in DeShon and Alexander (1996), Dretzke et al., and Overton (2001). Also, a related Z procedure proposed in Welch (1938) is presented in Appendix B for the sake of completeness and convenient reference.

WLS

Under the notion of heterogeneous error variance $\sigma_1^2 \neq \sigma_2^2$, Overton (2001) considered a WLS analysis of the problem using multiple regression model in Equation 2. In general, the properties of WLS estimators differ from those of OLS estimators; the WLS approach is employed to correct for heteroscedasticity, in that error variance changes as a function of covariate variables. For example, see Kutner, Nachtsheim, Neter, and Li (2005, sections 11.1 and 18.4). Moreover, WLS is a special case of generalized least squares, in which the error terms not only may have different variances but may also be correlated in pairs. However, the situation of heterogeneity of group error variances considered here requires only a single weight for each group. As reported in Overton (2001, p. 222), the WLS-based and OLS-based analyses of moderated multiple regression model (Equation 2) yield identical coefficient estimators for β_{20}, β_{21}, δ_0, and δ_1, but differ in their respective variance estimators. However, Overton did not provide the specific analytic formulations for the WLS estimator $\hat{\delta}_{W1}$ and the corresponding estimated variance $\hat{V}(\hat{\delta}_{W1})$. More importantly, the exact formulation of the test procedure was not given; instead, only numerical results were presented in Overton. Although empirical investigations are useful in assessing the properties of the competing methods, it is of pedagogical interest to see how different the WLS estimator is from other procedures that have been used in many applications. As expected, the WLS test statistic T_W for the test of H_0: $\delta_1 = 0$ can be expressed as

$$T_W = \frac{\hat{\delta}_{W1}}{\{\hat{V}(\hat{\delta}_{W1})\}^{1/2}}.$$

Even though it appears to be correctly specified, the general form does not provide much information about the theoretical implications of T_W. It follows from the matrix representation and manipulation, in Equation C7 of Appendix C, that the WLS estimator of δ_1 is $\hat{\delta}_{W1} = \hat{\beta}_{11} - \hat{\beta}_{21}$, regardless of the selection of relative weights. However, the estimated variance $\hat{V}(\hat{\delta}_{W1})$ varies with the designated weights. In particular, the WLS procedure of Overton (2001) employs the following weights:

$$\omega_i = \left\{\frac{N_i - 4}{SSE_i}\right\}^{1/2} = \left\{\frac{N_i - 4}{N_i - 2}\cdot\frac{1}{\hat{\sigma}_i^2}\right\}^{1/2}, \; i = 1 \text{ and } 2.$$

It follows from Equation C14 of Appendix C that the WLS test statistic for H_0: $\delta_1 = 0$ is

$$T_\omega = \frac{\hat{\beta}_{11} - \hat{\beta}_{21}}{\hat{V}_\omega^{1/2}}, \tag{6}$$

where $\hat{V}_\omega$ is defined in Equation C13. Also, T_ω has the approximate t distribution

$$T_\omega \stackrel{\cdot}{\sim} t(dfw), \tag{7}$$

where $dfw = N - 4$. The null hypothesis is rejected at the significance level α, if $|T_\omega| > t_{dfw,\alpha/2}$. Accordingly, the WLS procedure of Overton (2001) is the square of T_ω or F_ω ($= T_\omega^2$) in our notation.

An alternative modified WLS (MWLS) method is also examined in Overton (2001). Equations C9, C10, and C11 of Appendix C provide details about the derivation of the MWLS T_m method. It is important to emphasize that WLS estimators of $\delta_1 = \beta_{11} - \beta_{21}$ are identical for any proper selection of weights; however, the estimated variances are generally different, as are the variance estimates $\hat{V}_m$ and $\hat{V}_\omega$ in Equations C10 and C13 of the two WLS-based test statistics T_m and T_ω, respectively. Nonetheless, the only exception occurs in the special circumstance of balanced group sizes $N_1 = N_2$ that $\hat{V}_m = \hat{V}_\omega$, and T_m and T_ω with the same referred t distribution $t(dfw)$. Although this distinguishing property between two WLS criteria was not mentioned in Overton, this phenomenon was already shown in the simulation results of Type I error rate and power reported for the two tests in Tables 1 and 4 of Overton, respectively, for the balanced group size $N_1 = N_2 = 50$. However, Overton concluded that the MWLS $F_m = T_m^2$ test does not perform as well as the WLS test $F_\omega = T_\omega^2$, according to his extensive Monte Carlo investigations across a wide range of model configurations.

In contrast to the WLS analysis, the OLS regression analysis of the model in Equation 2 is straightforward. As shown in Equation C15 of Appendix C, the OLS test statistic for H_0: $\delta_1 = 0$ is

$$T_O = \frac{\hat{\beta}_{11} - \hat{\beta}_{21}}{\{\hat{\sigma}_O^2(1/SSX_1 + 1/SSX_2)\}^{1/2}},$$

where $\hat{\sigma}_O^2 = SSE_O/(N - 4)$ and $SSE_O = SSE_1 + SSE_2$. The null hypothesis is rejected at the significance level α if $|T_O| > t_{N-4,\alpha/2}$. It is well known that the pattern of test T_O parallels exactly what is known about the usual two-sample pooled-variance t test under homogeneous group error variances. As evidenced in the simulation studies of Overton (2001), the test T_O is outperformed by the WLS-based tests T_m and T_ω and Welch's V test under the condition of heterogeneous group error variance.

Distinction Between Welch and WLS Procedures

It is notable from the methodical presentations in the two preceding subsections that the Welch's procedures and the WLS methods are apparently developed from different perspectives. It is not particularly surprising to see that two distinct principles lead to substantially different formulations and properties for the developed methods. However, there are important similarities and differences between the methodological formulations of Welch's V and WLS-based T_m, and Welch's Z and WLS-based T_ω. The following discussions derive the relevant results to show the closer relation of these renowned tests.

First, it follows from the resulting expressions (Equations 3 and C11) that, in fact, the two statistics of Welch's V and the WLS-based T_m are identical. Therefore, it is worthwhile to note that Welch's V statistic accommodates implicitly the notion of WLS. However, the respective approximate t distributions of the two tests are different. Actually, it can be shown that this phenomenon also exists in the framework of the Behrens–Fisher problem for comparing two normal means under heterogeneous variance assumption; specifically, the corresponding degrees of freedom for the referred t distribution of V and T_m in Equations 4 and C11 are $\hat{\nu}$ and $dfw = N - 4$, respectively. Note that the degrees of freedom $\hat{\nu}$ of V is bounded between the minimum of ($N_1 - 2$, $N_2 - 2$) and $N - 4$. Therefore, the associated critical values of $t_{\hat{\nu},\alpha/2}$ and $t_{dfw,\alpha/2}$ are in the order of $t_{dfw,\alpha/2} \le t_{\hat{\nu},\alpha/2}$. See Ghosh (1973) for the monotonicity properties of the family of t distribution. Hence, the observed significance level or p value of the Welch's V test is always greater than or equal to that of the WLS T_m test; in other words, the WLS test is more liberal than the Welch approach is, in the sense that the WLS test tends to reject the null hypothesis more often than the Welch method does. Correspondingly, the achieved significance level and power of the Welch method never exceed those of the WLS procedure. Interestingly, these characteristics of the Welch and WLS approaches were not addressed in Overton (2001), yet the simulation studies in Tables 1–5 of Overton exemplified these features between the two approaches under the notations of F^* and MWLS for V and T_m, respectively.

Additionally, the prescribed Welch's Z test of Equation B2 and WLS-based T_ω procedure in Equation 7 are entangled in formulation and approximate distribution. In view of the distinctive form of the estimated variance $\hat{V}_\omega$ given in Equation C13, the statistic T_ω of Equations 6 and C14 is in relation to the Z statistic in Equation B1 through $d \cdot T_\omega = Z$, where $d = \{(N - 8)/(N - 4)\}^{1/2}$. Hence, the approximate distribution of $T_\omega \mathrel{\dot\sim} t(dfw)$ defined in Equation 7 can be rewritten as

$$d \cdot T_\omega = Z \mathrel{\dot\sim} d \cdot t(dfw).$$

On the contrary, Welch's approximation for Z is $Z \mathrel{\dot\sim} c \cdot t(f^*)$, as presented in Equation B2. Consequently, Welch's Z approach and the specific transformation $d \cdot T_\omega$ of the WLS-based T_ω method belong to a family of (approximate) distributions, each of which is approximated by $k \cdot t(f)$, where, for Z, $k = c$ and $f = f^*$, and for $d \cdot T_\omega$, $k = d$ and $f = dfw$. Note that the approximate $c \cdot t(f^*)$ distribution of Welch's Z is optimized in terms of the scalar multiplier c and degrees of freedom f^*. Therefore, the approximate distribution in terms of $d \cdot t(dfw)$ for the transformed statistic $d \cdot T_\omega$ or Z does not possess the optimality property and is less adequate under the moment-matching criterion of Welch (1938). Later, the differences between the WLS T_ω and Welch's V methods will be further examined and reinforced in the simulation study.

Simulation Study

On the basis of the present results, the WLS-based T_ω and Welch's V methods emerge as the prominent and representative test procedures of the two distinct WLS (error variance heterogeneity neutralization) and Welch (approximate degrees of freedom) methodologies. To further help clarify similarities and differences for the competing T_ω and V methods, simulation investigations are performed to examine their numerical performance. For ease of exposition, the two test procedures will be referred to in the remainder of this article as the WLS and Welch approaches, respectively.

It was concluded in Overton (2001) that the behavior in controlling the Type I error rate for the WLS method is not as spectacular as the Welch procedure and other formulas. On the other hand, the WLS test is virtually identical to the Welch procedure in its ability to detect a true interaction effect. However, the conditions in which the Welch and WLS procedures incur the most discrepancy were not identified. Hence, we performed an extensive replication of Overton's simulation study to reevaluate the empirical Type I error rate for the detection of interaction effects. Throughout the numerical study, the nominal Type I error rate is set as $\alpha = .05$. The estimates of the true Type I error rate associated with given sample size and model configurations are computed through Monte Carlo simulation of 10,000 independent data sets. For each replicate, N_1 and N_2 values of predictor X_1 and X_2 are generated from the designated independent normal distribution $N(0, \sigma_{X1}^2)$ and $N(0, \sigma_{X2}^2)$, respectively. These values in turn determine the respective mean responses for generating N_1 and N_2 values of normal outcomes Y_1 and Y_2 for the two underlying regression models with error variance σ_1^2 and σ_2^2, as defined in Equation 1. Then the two test statistics T_ω and V are computed. Accordingly, the simulated Type I error rate is the proportion of the 10,000 replicates, whose values of $|T_\omega|$ and $|V|$ exceed the critical values $t_{dfw,.025}$ and $t_{\hat{\nu},.025}$ for the WLS and Welch procedures, respectively. As in Overton (2001, p. 223), the three harmonic group sample size means are 20, 50, and 100, and the ratio of sample sizes varies from 1:1 to 1:2 to 1:5. The resulting sample sizes (N_1, N_2) combinations are (20, 20), (15, 30), and (12, 60) for the harmonic mean of 20; (50, 50), (38, 75), and (30, 150) for the harmonic mean of 50; and (100, 100), (75, 150), and (60, 300) for the harmonic mean of 100. For each of the three harmonic means of 20, 50, and 100, a total of 88 model settings are summarized in three tables, according to the combined configurations of sample size allocation (N_1 and N_2), predictor standard deviation (*SD*) ratio (σ_{X1}/σ_{X2}), and error *SD* ratio (σ_1/σ_2). Overall, the numerical results are summarized in a total of nine tables. Although the results fit in with the general conclusions of Overton, some important and distinctive situations could still be of practical interest, in the sense that the two contending procedures have the more obvious potential of yielding different conclusions. Space limitations preclude reporting results for all situations. To exemplify the critical and subtle discrepancy between the two approaches, only the tables in which the harmonic mean sample size is 20 are presented. Tables 1–3 contain the simulated Type I error rates for sample sizes (N_1, N_2) = (20, 20), (15, 30), and (12, 60), respectively. The full set of simulation results is available upon request.

It was noted in Overton (2001) that the Type I error rates of the WLS test ranged from .044 to .060, and 91% of the 264 condition error rates were in the narrow .045–.055 range. According to our simulations, the simulated Type I error rates of WLS had a similar range of .0444 to .0606, where the value 0.0606 is associated with Condition 31 in Table 2. Nonetheless, only 85.23% (225 cases) of the 264 conditions were in the interval of .045–.055. In particular, there are 8, 13, and 7 cases in Tables 1–3, respectively, with large deviation not inside the prescribed range. Hence, the corresponding percentage of the simulated Type I error rates that is within the range of .045–.055 in Tables 1–3 is as low as 68.18% (60 out of the 88 conditions). Moreover, for the 28 cases outside the range of .045–.055, only a single value, .0445 (Condition 15 in Table 3), fell below .045. It appears that the WLS method tends to give less precise and positively biased Type I error for comparatively small sample sizes. On the contrary, the Welch procedure yielded a range of .0447 to .0555 with 98.49% (260/264) between .045 and .055. Incidentally, it is noteworthy from Tables 1–3 that Condition 35 in Table 2 is the only case of the Welch test to incur a large deviation (.0051) not within the bound of .005. The remarkably accurate results of the Welch procedure presented here are consistent with the findings in Overton (2001, p. 224). Hence, it is clear that the Welch procedure has the important advantage over the WLS test of accurate control of Type I error rate, especially for small samples. Moreover, since the considered distri-

Table 1
Simulated Type I Error Rates of WLS and Welch Procedures When Sample Sizes N_1 = 20 and N_2 = 20 (α = .05)

Condition	σ_{X1}/σ_{X2}	σ_1/σ_2	WLS	Welch	Difference WLS − .05	Difference Welch − .05
1	1/1	1/1	.0532	.0526	.0032	.0026
2	1/1	1/2	.0528	.0513	.0028	.0013
3	1/1	1/4	.0523	.0465	.0023	−.0035
4	1/1	1/8	.0552	.0485	.0052	−.0015
5	1/2	1/1	.0491	.0466	−.0009	−.0034
6	1/2	1/2	.0500	.0491	.0000	−.0009
7	1/2	1/4	.0513	.0482	.0013	−.0018
8	1/2	1/8	.0542	.0494	.0042	−.0006
9	1/2	2/1	.0561	.0504	.0061	.0004
10	1/2	4/1	.0570	.0489	.0070	−.0011
11	1/2	8/1	.0565	.0492	.0065	−.0008
12	1/4	1/1	.0574	.0511	.0074	.0011
13	1/4	1/2	.0529	.0504	.0029	.0004
14	1/4	1/4	.0531	.0529	.0031	.0029
15	1/4	1/8	.0505	.0482	.0005	−.0018
16	1/4	2/1	.0580	.0502	.0080	.0002
17	1/4	4/1	.0591	.0519	.0091	.0019
18	1/4	8/1	.0589	.0495	.0089	−.0005

Table 2
Simulated Type I Error Rates of WLS and Welch Procedures When Sample Sizes N_1 = 15 and N_2 = 30 (α = .05)

Condition	σ_{X1}/σ_{X2}	σ_1/σ_2	WLS	Welch	Difference WLS − .05	Difference Welch − .05
1	1/1	1/1	.0501	.0495	.0001	−.0005
2	1/1	1/2	.0512	.0510	.0012	.0010
3	1/1	1/4	.0510	.0478	.0010	−.0022
4	1/1	1/8	.0537	.0478	.0037	−.0022
5	1/1	2/1	.0553	.0521	.0053	.0021
6	1/1	4/1	.0567	.0509	.0067	.0009
7	1/1	8/1	.0588	.0517	.0088	.0017
8	1/2	1/1	.0558	.0521	.0058	.0021
9	1/2	1/2	.0512	.0506	.0012	.0006
10	1/2	1/4	.0488	.0490	−.0012	−.0010
11	1/2	1/8	.0522	.0485	.0022	−.0015
12	1/2	2/1	.0569	.0509	.0069	.0009
13	1/2	4/1	.0546	.0487	.0046	−.0013
14	1/2	8/1	.0563	.0501	.0063	.0001
15	2/1	1/1	.0468	.0471	−.0032	−.0029
16	2/1	1/2	.0505	.0470	.0005	−.0030
17	2/1	1/4	.0534	.0487	.0034	−.0013
18	2/1	1/8	.0552	.0494	.0052	−.0006
19	2/1	2/1	.0532	.0529	.0032	.0029
20	2/1	4/1	.0535	.0497	.0035	−.0003
21	2/1	8/1	.0586	.0536	.0086	.0036
22	1/4	1/1	.0588	.0532	.0088	.0032
23	1/4	1/2	.0523	.0494	.0023	−.0006
24	1/4	1/4	.0475	.0471	−.0025	−.0029
25	1/4	1/8	.0484	.0486	−.0016	−.0014
26	1/4	2/1	.0519	.0466	.0019	−.0034
27	1/4	4/1	.0593	.0533	.0093	.0033
28	1/4	8/1	.0573	.0508	.0073	.0008
29	4/1	1/1	.0543	.0513	.0043	.0013
30	4/1	1/2	.0532	.0479	.0032	−.0021
31	4/1	1/4	.0606	.0530	.0106	.0030
32	4/1	1/8	.0549	.0495	.0049	−.0005
33	4/1	2/1	.0461	.0460	−.0039	−.0040
34	4/1	4/1	.0499	.0496	−.0001	−.0004
35	4/1	8/1	.0581	.0551	.0081	.0051

Table 3
Simulated Type I Error Rates of WLS and Welch Procedures When Sample Sizes $N_1 = 12$ and $N_2 = 60$ ($\alpha = .05$)

Condition	σ_{X1}/σ_{X2}	σ_1/σ_2	WLS	Welch	Difference WLS − .05	Difference Welch − .05
1	1/1	1/1	.0528	.0529	.0028	.0029
2	1/1	1/2	.0486	.0512	−.0014	.0012
3	1/1	1/4	.0468	.0486	−.0032	−.0014
4	1/1	1/8	.0505	.0498	.0005	−.0002
5	1/1	2/1	.0564	.0519	.0064	.0019
6	1/1	4/1	.0542	.0504	.0042	.0004
7	1/1	8/1	.0566	.0504	.0066	.0004
8	1/2	1/1	.0517	.0480	.0017	−.0020
9	1/2	1/2	.0514	.0515	.0014	.0015
10	1/2	1/4	.0474	.0518	−.0026	.0018
11	1/2	1/8	.0488	.0505	−.0012	.0005
12	1/2	2/1	.0543	.0488	.0043	−.0012
13	1/2	4/1	.0532	.0482	.0032	−.0018
14	1/2	8/1	.0561	.0500	.0061	.0000
15	2/1	1/1	.0445	.0484	−.0055	−.0016
16	2/1	1/2	.0526	.0548	.0026	.0048
17	2/1	1/4	.0506	.0496	.0006	−.0004
18	2/1	1/8	.0493	.0466	−.0007	−.0034
19	2/1	2/1	.0510	.0512	.0010	.0012
20	2/1	4/1	.0523	.0483	.0023	−.0017
21	2/1	8/1	.0566	.0512	.0066	.0012
22	1/4	1/1	.0589	.0540	.0089	.0040
23	1/4	1/2	.0548	.0511	.0048	.0011
24	1/4	1/4	.0483	.0492	−.0017	−.0008
25	1/4	1/8	.0452	.0471	−.0048	−.0029
26	1/4	2/1	.0524	.0481	.0024	−.0019
27	1/4	4/1	.0544	.0488	.0044	−.0012
28	1/4	8/1	.0585	.0528	.0085	.0028
29	4/1	1/1	.0459	.0487	−.0041	−.0013
30	4/1	1/2	.0515	.0508	.0015	.0008
31	4/1	1/4	.0508	.0489	.0008	−.0011
32	4/1	1/8	.0550	.0523	.0050	.0023
33	4/1	2/1	.0453	.0493	−.0047	−.0007
34	4/1	4/1	.0531	.0534	.0031	.0034
35	4/1	8/1	.0549	.0511	.0049	.0011

butions in Equations 7 and 4 for the WLS and Welch tests are approximations, care needs to be taken in interpreting the implications of their results in the simulated power of Overton. It is important to note that the magnitude of empirical power of both procedures depends mostly on the accuracy of their respective approximate critical values. Therefore, it should be taken into consideration that the simulated power of the WLS test may be attained at the cost of an inordinate or unstable Type I error rate in the same conditions.

From a practical standpoint of providing a generally useful and versatile solution without specifically confining itself to any particular settings, the failure to give an accurate Type I error rate is one obvious limitation of the WLS test. Although not all study designs are planned with moderate or small sample sizes, it is understandable that some intrinsically original or special research would accommodate larger numbers of participants with difficulty. In this respect, it is essential for researchers to have a reliable procedure for detecting the moderating effects over all sample size situations one might encounter in applied work. The soundness of Welch's approach and its primitive form have received critical acclaims from numerous researchers. The newly recognized trait of WLS enhances the incredible versatility of Welch's methodology. More importantly, the overall adequate performance in Type I error rates for detecting interaction effects fortifies the distinct advantage of the Welch procedure over the WLS method. As suggested by a referee, we also investigated the behavior of the Welch procedure under imperfect conditions. Accordingly, Monte Carlo simulations have been performed for the MMR analysis with two types of nonnormal errors—namely, gamma and uniform distributions. Since this issue is not the primary focus of the present article, the details are not given here. However, the performance seems completely acceptable regarding the robustness of the Welch procedure against mild departures from normal error assumption. Finally, the computational aspect for assessing moderating effects in the context of an MMR example will be described in the next section.

Example

In addition to the statistical performance in Type I error rate and power, Overton (2001) emphasized the practical importance of computational requirement

Table 4
The Hypothetical Values of Response Variable Y and Predictor Variable X

Y	X	Y	X	Y	X	Y	X	Y	X
				Group 1 ($Z = 1$)					
0.8427	−0.6411	0.7367	−0.4710	1.3724	−0.0829	1.0208	−0.9724	1.0173	1.9320
−0.2700	−0.5577	−1.1647	0.5300	−1.5785	−0.7436	−0.5066	−0.0115	−0.3525	0.5926
0.5504	0.1367	1.7815	2.0942						
				Group 2 ($Z = 0$)					
20.6221	5.8494	0.0773	−4.6409	7.2835	4.0269	−4.8243	2.3101	−8.1438	−5.7033
8.1696	1.0393	0.3258	3.3170	−11.5477	−5.2614	−3.7603	−0.3760	11.9974	7.8449
−10.0975	−5.0683	−1.3754	−1.4049	−9.6727	−4.3949	−18.7952	−3.3753	4.8187	−3.8163
17.8028	5.2447	1.5166	−1.7987	−18.1552	0.2492	16.7311	−0.5163	−6.6831	−0.6478
4.1600	−4.0378	−20.5484	−4.2237	−3.4620	−5.1611	5.8139	11.1631	−2.2820	7.1399
−5.3675	1.6372	3.1738	4.3391	2.8675	−3.2008	−0.0571	−2.4078	−5.4507	1.3559
−5.1712	4.8134	17.2321	2.4378	−2.5326	0.3754	−11.2807	−3.6007	−0.4768	2.7956
−9.5239	−5.8884	−5.5169	−2.3669	11.5640	3.7946	4.0643	1.2667	12.5105	1.5345
11.6301	3.3982	−1.6980	−2.0588	−9.9924	−0.7673	4.4596	7.6824	10.2650	3.2056
−7.4692	0.3997	−4.5522	4.5721	−3.7881	−3.0503	10.4150	8.9247	5.1856	4.9535
−0.6495	−4.8053	−4.6085	3.9554	1.7367	−0.4456	6.5870	2.2837	−5.1916	−0.4220
−5.2506	−0.1328	10.1324	0.4992	−5.8459	0.7123	1.3268	3.2873	−3.5859	−3.8313

and program availability of the competing methods. It was shown in Overton that the WLS-based moderated multiple regression analysis can be readily conducted with standard SAS procedures. For the practical purpose of expediting the application of the suggested Welch method, the corresponding SAS (SAS Institute, 2008) and R (R Development Core Team, 2008) computing algorithms for performing the Welch procedure are developed. In the process, we also show numerical evidence below that the procedures for testing moderating effects can have an important impact on the results, conclusions, and, ultimately, the theory that involves moderated relationships. To facilitate the following illustration in the context of MMR analysis, it is constructive to consider briefly whether the exemplifying aim of the numerical study is to determine the changes, as a function of gender group membership (Z), in the relationship between the employer's job performance (Y) and preemployment test score (X). As Aguinis and Pierce (1998) pointed out, one of the most typical situations in organizational study is when the group with the larger sample size is associated with the larger error variance. To exemplify this notable circumstance of practical importance, the two groups of values (Y, X) listed in Table 4 represent random samples generated from the underlying model configurations with $(\sigma_{X1}, \sigma_{X2}) = (1, 4)$, $(\sigma_1, \sigma_2) = (1, 8)$, $(\beta_{11}, \beta_{21}) = (0.15, 1.4223)$ for group sample sizes $(N_1, N_2) = (12, 60)$.

Accordingly, the value of the Welch test statistic defined in Equation 3 is $V = -2.0740$ with degrees of freedom $\hat{\nu} = 24.7708$ and p value $= .0486$, whereas the WLS method yields the outcomes of $T_\omega = -1.9786$, $dfw = 68$, and p value $= .0519$ for this data set. Conceivably, although the two values of the Welch and WLS test statistics are only slightly different, the embedded subtle approximation features lead to dissimilar conclusions on the basis of significant level $\alpha = .05$. Moreover, the OLS test results in $T_O = -0.3812$, with p value $= .7043$. Thus, the extraordinarily large p value of OLS gives an apparently disparate and implausible outcome, compared with the Welch procedure. This is not a particularly surprising result, however, in view of the existing numerical findings that the Type I error rate of the ordinary F test is excessively conservative when the group with the larger sample size is associated with the larger error variance.

In general, the established notion of the Welch procedure in adjusting the degrees of freedom for the approximate distribution has resulted in wide acceptance in the literature. The SAS and R programs for the implementation of the Welch procedure presented in Appendixes D and E are available to interested researchers upon request. Users can easily identify the statements with the self-contained exposition, and it only requires a slight modification of the program to accommodate their own data specifications.

Conclusions

Several previous investigations have shown that the standard MMR F test is not robust to the heterogeneity of error variance for evaluating regression slope differences across groups. In addition to the existing attempts, the WLS procedure of Overton (2001) provides an attractive alternative to mitigate the impact of violating the assumption of homogeneous error variance on conclusions of testing the hypotheses regarding the interaction between a dichotomous moderator variable and a continuous predictor variable. For pedagogic reasons, one must have a thorough understanding of the fundamental details of the methodology, and how the technique improves upon existing approaches for MMR analysis, before the theoretical idea can finally be considered appropriate for making sound application. This article elucidates the similarities and differences between the Welch and WLS methods through rigorous analytical presentations and numerical assessments. In particular, it shows how the Welch statistics have exactly the same or similar expressions as do the WLS-based MMR statistics. Therefore, from the methodological point of view, Welch's procedure exhorts the same tactic as the WLS method for tackling the problem of error variance heterogeneity; in other words, the prevailing Welch approach implicitly possesses the same tempting property that distinguishes the WLS method from other available techniques. However, the resulting testing procedures differ in their adjustments of the degrees of freedom for respective distributional approximations. Furthermore, the primitive Welch approaches for comparing means of two and more groups under heterogeneity of variance have been widely discussed in standard texts of statistical methods in psychology and business (see, e.g., Berenson, Krehbiel, & Levine, 2006; Howell, 2007). Although there is some concern about the application of the Welch procedure for four or more groups (DeShon & Alexander, 1996), this research has been confined to the case of two groups. With the underlying WLS characteristic, accurate performance and computational ease, it is prudent to recommend the extended Welch's approximate t procedure for detecting the moderated effects of dichotomous moderators in MMR.

AUTHOR NOTE

G.S. thanks the two anonymous referees for their valuable comments on earlier drafts of the article. This research was partially supported by the National Science Council. Correspondence concerning this article should be addressed to G. Shieh, Department of Management Science, National Chiao Tung University, Hsinchu, Taiwan 30050 (e-mail: gwshieh@mail.nctu.edu.tw).

REFERENCES

Aguinis, H. (2004). *Regression analysis for categorical moderators*. New York: Guilford.

Aguinis, H., Petersen, S. A., & Pierce, C. A. (1999). Appraisal of the homogeneity of error variance assumption and alternatives for multiple regression for estimating moderated effects of categorical variables. *Organizational Research Methods*, **2**, 315-339.

Aguinis, H., & Pierce, C. A. (1998). Heterogeneity of error variance and the assessment of moderating effects of categorical variables: A conceptual review. *Organizational Research Methods*, **1**, 296-314.

Aguinis, H., & Stone-Romero, E. F. (1997). Methodological artifacts in moderated multiple regression and their effects on statistical power. *Journal of Applied Psychology*, **82**, 192-206.

ALEXANDER, R. A., & DESHON, R. P. (1994). The effect of error variance heterogeneity on the power of tests for regression slope differences. *Psychological Bulletin*, **115**, 308-314.

ALEXANDER, R. A., & GOVERN, D. M. (1994). A new and simpler approximation for ANOVA under variance heterogeneity. *Journal of Educational Statistics*, **19**, 91-101.

BERENSON, M. L., KREHBIEL, T. C., & LEVINE, D. M. (2006). *Basic business statistics: Concepts and applications* (10th ed.). Upper Saddle River, NJ: Pearson Education.

BEST, D. J., & RAYNER, J. C. W. (1987). Welch's approximate solution for the Behrens–Fisher problem. *Technometrics*, **29**, 205-210.

DAVENPORT, J. M., & WEBSTER, J. T. (1975). The Behrens–Fisher problem, an old solution revisited. *Metrika*, **22**, 47-54.

DESHON, R. P., & ALEXANDER, R. A. (1994). A generalization of James's second-order approximation to the test for regression slope equality. *Educational & Psychological Measurement*, **54**, 328-335.

DESHON, R. P., & ALEXANDER, R. A. (1996). Alternative procedures for testing regression slope homogeneity when group error variances are unequal. *Psychological Methods*, **1**, 261-277.

DRETZKE, B. J., LEVIN, J. R., & SERLIN, R. C. (1982). Testing for regression homogeneity under variance heterogeneity. *Psychological Bulletin*, **91**, 376-383.

FENSTAD, G. U. (1983). A comparison between the *U* and *V* tests in the Behrens–Fisher problem. *Biometrika*, **70**, 300-302.

GHOSH, B. K. (1973). Some monotonicity theorems for χ^2, *F* and *t* distributions with applications. *Journal of the Royal Statistical Society B*, **35**, 480-492.

HOWELL, D. C. (2007). *Statistical methods for psychology* (6th ed.). Pacific Grove, CA: Duxbury.

JAMES, G. S. (1951). The comparison of several groups of observations when the ratios of the population variances are unknown. *Biometrika*, **38**, 324-329.

NEL, D. G., VAN DER MERWE, C. A., & MOSER, B. K. (1990). The exact distribution of the univariate and multivariate Behrens–Fisher statistics with a comparison of several solutions in the univariate case. *Communications in Statistics: Theory & Methods*, **19**, 279-298.

NETER, J., KUTNER, M. H., NACHTSHEIM, C. J., & LI, W. (2005). *Applied linear statistical models* (5th ed.). Chicago: McGraw-Hill/Irwin.

OVERTON, R. C. (2001). Moderated multiple regression for interactions involving categorical variables: A statistical control for heterogeneous variance across two groups. *Psychological Methods*, **6**, 218-233.

PAUL, S. R., BEST, D. J., & RAYNER, J. C. W. (1992). Comment on Best and Rayner (1987). *Technometrics*, **34**, 249-250.

R DEVELOPMENT CORE TEAM (2008). *R: A language and environment for statistical computing* [Computer software and manual]. Retrieved from www.r-project.org.

RENCHER, A. C. (2000). *Linear models in statistics*. New York: Wiley.

SAS INSTITUTE (2008). *SAS/STAT 9.2 user's guide*. Cary, NC: SAS Institute Inc.

SCARIANO, S. M., & DAVENPORT, J. M. (1986). A four-moment approach and other practical solutions to the Behrens–Fisher problem. *Communications in Statistics: Theory & Methods*, **15**, 1467-1505.

SCHEFFÉ, H. (1970). Practical solutions of the Behrens–Fisher problem. *Journal of the American Statistical Association*, **65**, 1501-1508.

STONE-ROMERO, E. F., ALLIGER, G. M., & AGUINIS, H. (1994). Type II error problems in the use of moderated multiple regression for the detection of moderating effects of dichotomous variables. *Journal of Management*, **20**, 167-178.

WELCH, B. L. (1938). The significance of the difference between two means when the population variances are unequal. *Biometrika*, **29**, 350-362.

WELCH, B. L. (1951). On the comparison of several mean values: An alternative approach. *Biometrika*, **38**, 330-336.

SUPPLEMENTAL MATERIALS

Descriptions of the analytical procedures used and SAS and R code to implement them may be downloaded as supplemental materials for this article from brm.psychonomic-journals.org/content/supplemental.

APPENDIX A
The Subgroup OLS Analysis

Let $\mathbf{Y}_1$ and $\mathbf{Y}_2$ be $N_1 \times 1$ and $N_2 \times 1$ column vectors of Y_{1i}s and Y_{2j}s. Denote $\mathbf{X}_{D1} = [\mathbf{1}_{N1}, \mathbf{X}_1]$ and $\mathbf{X}_{D2} = [\mathbf{1}_{N2}, \mathbf{X}_2]$, where $\mathbf{1}_{N1}$ and $\mathbf{1}_{N2}$ are $N_1 \times 1$ and $N_2 \times 1$ column vectors of 1s, and $\mathbf{X}_1$ and $\mathbf{X}_2$ are the $N_1 \times 1$ and $N_2 \times 1$ column vectors of X_{1i}s and X_{2j}s. Then, the matrix formulations of the two models in Equation 1 are $\mathbf{Y}_1 = \mathbf{X}_{D1}\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}_1$ and $\mathbf{Y}_2 = \mathbf{X}_{D2}\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}_2$, respectively, where $\boldsymbol{\beta}_1 = [\beta_{10}, \beta_{11}]^{\mathrm{T}}$ and $\boldsymbol{\beta}_2 = [\beta_{20}, \beta_{21}]^{\mathrm{T}}$. Consequently, the separate OLS estimators are

$$\hat{\boldsymbol{\beta}}_1 = \begin{bmatrix} \hat{\beta}_{10} \\ \hat{\beta}_{11} \end{bmatrix} = (\mathbf{X}_{D1}^{\mathrm{T}}\mathbf{X}_{D1})^{-1}\mathbf{X}_{D1}^{\mathrm{T}}\mathbf{Y}_1 \text{ and } \hat{\boldsymbol{\beta}}_2 = \begin{bmatrix} \hat{\beta}_{20} \\ \hat{\beta}_{21} \end{bmatrix} = (\mathbf{X}_{D2}^{\mathrm{T}}\mathbf{X}_{D2})^{-1}\mathbf{X}_{D2}^{\mathrm{T}}\mathbf{Y}_2. \quad \text{(A1)}$$

Also, the variance and covariance matrices are

$$\mathrm{Cov}(\hat{\boldsymbol{\beta}}_1) = \sigma_1^2(\mathbf{X}_{D1}^{\mathrm{T}}\mathbf{X}_{D1})^{-1} \text{ and } \mathrm{Cov}(\hat{\boldsymbol{\beta}}_2) = \sigma_2^2(\mathbf{X}_{D2}^{\mathrm{T}}\mathbf{X}_{D2})^{-1}, \quad \text{(A2)}$$

respectively.

APPENDIX B
Welch's *Z* Procedure

One of the alternative methods suggested in Welch (1938, p. 360) for the Behrens–Fisher problem is the *Z* statistic (see Fenstad, 1983; Best & Rayner, 1987; and Paul, Best, & Rayner, 1992, for further details). Although it was not explicitly stated, the notion of the *Z* test for use in the Behrens–Fisher situation can be easily generalized for the comparison of regression coefficients as well. Following Welch's (1938) derivation, and that of Paul et al. (1992), the extended *Z* statistic has the form

$$Z = \frac{\hat{\beta}_{11} - \hat{\beta}_{21}}{[(N_1 - 2)\hat{\sigma}_1^2 / \{(N_1 - 4)SSX_1\} + (N_2 - 2)\hat{\sigma}_2^2 / \{(N_2 - 4)SSX_2\}]^{1/2}}. \quad \text{(B1)}$$

It follows from Welch (1938) under the null hypothesis H_0: $\beta_{11} = \beta_{21}$ that the distribution of *Z* can be approximated by a scalar multiplier of *t* distribution

$$Z \mathrel{\dot{\sim}} c \cdot t(f^*), \quad \text{(B2)}$$

APPENDIX B (Continued)

where

$$c = \left[\frac{\hat{\sigma}_1^2 / SSX_1 + \hat{\sigma}_2^2 / SSX_2}{(N_1 - 2)\hat{\sigma}_1^2 / \{(N_1 - 4)SSX_1\} + (N_2 - 2)\hat{\sigma}_2^2 / \{(N_2 - 4)SSX_2\}} \right]^{1/2},$$

and $t(f^*)$ is a t distribution with degrees of freedom

$$f^* = \frac{[(N_1 - 2)\hat{\sigma}_1^2 / \{(N_1 - 4)SSX_1\} + (N_2 - 2)\hat{\sigma}_2^2 / \{(N_2 - 4)SSX_2\}]^2}{(N_1 - 2)[\hat{\sigma}_1^2 / \{(N_1 - 4)SSX_1\}]^2 + (N_2 - 2)[\hat{\sigma}_2^2 / \{(N_2 - 4)SSX_2\}]^2}.$$

Although the Welch statistics of V in Equation 3 and Z just described in Equation B1 are equally efficient in terms of asymptotic relative efficiency, it is important to note that the finite sample comparisons of Best and Rayner (1987) and Paul et al. (1992) recommended the statistic V over the statistic Z for testing the hypothesis of equality of two normal means. In the case of comparison of regression coefficients, it is conceivable that the test procedure V in Equation 4 should possess the same advantage over the Z method in Equation B2.

APPENDIX C
The WLS Procedures

Consider the matrix formulation of the model in Equation 2,

$$\mathbf{Y} = \mathbf{X}_\mathrm{D}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{C1}$$

where $\boldsymbol{\beta} = [\beta_{20}, \beta_{21}, \delta_0, \delta_1]^\mathrm{T}$ is the 4×1 column vector of regression coefficients; $\mathbf{Y}$ and $\boldsymbol{\varepsilon}$ are the respective column vectors of Y_k and ε_k, $k = 1, \ldots, N$; and $\mathbf{X}_\mathrm{D} = [\mathbf{1}_N, \mathbf{X}, \mathbf{Z}, \mathbf{ZX}]$ is the $N \times 4$ design matrix with $\mathbf{1}_N$, $\mathbf{X}$, $\mathbf{Z}$, and $\mathbf{ZX}$, which are the column vectors of all 1s, X_ks, Z_ks, and $Z_k X_k$s, respectively. Under the variance heterogeneity assumption, the variance–covariance matrix of $\boldsymbol{\varepsilon}$ is $\mathrm{Cov}(\boldsymbol{\varepsilon}) = \mathrm{Diag}\{\sigma_2^2\mathbf{I}_{N_2}, \sigma_1^2\mathbf{I}_{N_1}\}$, an $N \times N$ diagonal matrix, where $\mathbf{I}_{N_1}$ and $\mathbf{I}_{N_2}$ are the identity matrixes of dimensions N_1 and N_2, respectively. The WLS approach modified the ordinary least squares method by applying an appropriate weight matrix to the model in Equation C1 as follows:

$$\mathbf{Y}^* = \mathbf{X}_\mathrm{D}^*\boldsymbol{\beta} + \boldsymbol{\varepsilon}^*,$$

where $\mathbf{Y}^* = \mathbf{WY}$, $\mathbf{X}_\mathrm{D}^* = \mathbf{WX}_\mathrm{D}$, $\boldsymbol{\varepsilon}^* = \mathbf{W}\boldsymbol{\varepsilon}$, and the weight matrix $\mathbf{W} = \mathrm{Diag}\{\mathbf{W}_2, \mathbf{W}_1\}$ is the $N \times N$ diagonal matrix with the first N_2 and the last N_1 diagonal elements equal to w_2 and w_1, where $\mathbf{W}_2 = w_2\mathbf{I}_{N_2}$ and $\mathbf{W}_1 = w_1\mathbf{I}_{N_1}$. Let $\mathrm{Var}(\varepsilon_k^*) = \sigma_\mathrm{w}^2$, then $\sigma_\mathrm{w}^2 = w_2^2\sigma_2^2$ for $k = 1, \ldots, N_2$ and $\sigma_\mathrm{w}^2 = w_1^2\sigma_1^2$ for $k = N_2 + 1, \ldots, N_2 + N_1$. The subsequent WLS analysis follows that of the regular OLS linear regression. Hence, the WLS estimator $\hat{\boldsymbol{\beta}}_\mathrm{W} = [\hat{\beta}_{\mathrm{W}20}, \hat{\beta}_{\mathrm{W}21}, \hat{\delta}_{\mathrm{W}0}, \hat{\delta}_{\mathrm{W}1}]^\mathrm{T}$ of $\boldsymbol{\beta}$ can be readily obtained as

$$\hat{\boldsymbol{\beta}}_\mathrm{W} = (\mathbf{X}_\mathrm{D}^{*\mathrm{T}}\mathbf{X}_\mathrm{D}^*)^{-1}\mathbf{X}_\mathrm{D}^{*\mathrm{T}}\mathbf{Y}^* = (\mathbf{X}_\mathrm{D}^\mathrm{T}\mathbf{W}^2\mathbf{X}_\mathrm{D})^{-1}\mathbf{X}_\mathrm{D}^\mathrm{T}\mathbf{W}^2\mathbf{Y}. \tag{C2}$$

The variance–covariance matrix of $\hat{\boldsymbol{\beta}}_\mathrm{W}$ is $\mathbf{V}_\mathrm{W} = \mathrm{Cov}(\hat{\boldsymbol{\beta}}_\mathrm{W}) = \sigma_\mathrm{w}^2(\mathbf{X}_\mathrm{D}^{*\mathrm{T}}\mathbf{X}_\mathrm{D}^*)^{-1}$ and the corresponding natural estimator is

$$\hat{\mathbf{V}}_\mathrm{W} = \hat{\sigma}_\mathrm{w}^2(\mathbf{X}_\mathrm{D}^\mathrm{T}\mathbf{W}^2\mathbf{X}_\mathrm{D})^{-1}, \tag{C3}$$

where $\hat{\sigma}_\mathrm{w}^2 = SSE_\mathrm{W}/(N - 4)$ and the error sum of squares

$$SSE_\mathrm{W} = (\mathbf{Y}^* - \mathbf{X}_\mathrm{D}^*\hat{\boldsymbol{\beta}}_\mathrm{W})^\mathrm{T}(\mathbf{Y}^* - \mathbf{X}_\mathrm{D}^*\hat{\boldsymbol{\beta}}_\mathrm{W}) = (\mathbf{Y} - \mathbf{X}_\mathrm{D}\hat{\boldsymbol{\beta}}_\mathrm{W})^\mathrm{T}\mathbf{W}^2(\mathbf{Y} - \mathbf{X}_\mathrm{D}\hat{\boldsymbol{\beta}}_\mathrm{W}).$$

It follows that the hypothesis of H_0: $\delta_1 = 0$ or H_0: $\beta_{11} = \beta_{21}$ is therefore tested with the partial t statistic

$$T_\mathrm{W} = \frac{\hat{\delta}_{\mathrm{W}1}}{\left\{\hat{\mathrm{V}}(\hat{\delta}_{\mathrm{W}1})\right\}^{1/2}}, \tag{C4}$$

where $\hat{\delta}_{\mathrm{W}1}$ is the WLS estimator of δ_1 and $\hat{\mathrm{V}}(\hat{\delta}_{\mathrm{W}1})$ is the estimated variance of $\hat{\delta}_{\mathrm{W}1}$ and is the (4, 4) element of $\hat{\mathbf{V}}_\mathrm{W}$. The null hypothesis is rejected at the significance level α if $|T_\mathrm{W}| > t_{dfw,\alpha/2}$, where $dfw = N - 4$. Although the prescribed results are sufficient for the purpose of numerical computation for the selected weight matrix, it is not obvious from the general expressions exactly what the particular estimator $\hat{\delta}_{\mathrm{W}1}$ and the associated variance estimator $\hat{\mathrm{V}}(\hat{\delta}_{\mathrm{W}1})$ turn out to be. The following detailed presentation and illustration define the analytic results for the WLS approach and provide the connections of WLS with other related procedures.

Note that the overall response $\mathbf{Y}$ and the design matrix $\mathbf{X}_\mathrm{D}$ can be expressed as

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_2 \\ \mathbf{Y}_1 \end{bmatrix} \text{ and } \mathbf{X}_\mathrm{D} = \begin{bmatrix} \mathbf{X}_{\mathrm{D}2} & \mathbf{0}_{N_2 \times 2} \\ \mathbf{X}_{\mathrm{D}1} & \mathbf{X}_{\mathrm{D}1} \end{bmatrix},$$

APPENDIX C (Continued)

where $\mathbf{0}_{N_2\times 2}$ is an $N_2 \times 2$ matrix of 0s. It follows from the block diagonal property of $\mathbf{W}$ and the unique constant multiple of identity for the two weight matrices of $\mathbf{W}_1$ and $\mathbf{W}_2$ that

$$\mathbf{Y}^* = \begin{bmatrix} w_2\mathbf{Y}_2 \\ w_1\mathbf{Y}_1 \end{bmatrix}, \qquad \mathbf{X}_\mathrm{D}^* = \begin{bmatrix} w_2\mathbf{X}_{\mathrm{D2}} & \mathbf{0}_{N_2\times 2} \\ w_1\mathbf{X}_{\mathrm{D1}} & w_1\mathbf{X}_{\mathrm{D1}} \end{bmatrix},$$

$$\mathbf{X}_\mathrm{D}^{*\mathrm{T}}\mathbf{Y}^* = \begin{bmatrix} w_2^2\mathbf{X}_{\mathrm{D2}}^\mathrm{T}\mathbf{Y}_2 + w_1^2\mathbf{X}_{\mathrm{D1}}^\mathrm{T}\mathbf{Y}_1 \\ w_1^2\mathbf{X}_{\mathrm{D1}}^\mathrm{T}\mathbf{Y}_1 \end{bmatrix}$$

and

$$\left(\mathbf{X}_\mathrm{D}^{*\mathrm{T}}\mathbf{X}_\mathrm{D}^*\right)^{-1} = \begin{bmatrix} \left(w_2^2\mathbf{X}_{\mathrm{D2}}^\mathrm{T}\mathbf{X}_{\mathrm{D2}}\right)^{-1} & -\left(w_2^2\mathbf{X}_{\mathrm{D2}}^\mathrm{T}\mathbf{X}_{\mathrm{D2}}\right)^{-1} \\ -\left(w_2^2\mathbf{X}_{\mathrm{D2}}^\mathrm{T}\mathbf{X}_{\mathrm{D2}}\right)^{-1} & \left(w_1^2\mathbf{X}_{\mathrm{D1}}^\mathrm{T}\mathbf{X}_{\mathrm{D1}}\right)^{-1} + \left(w_2^2\mathbf{X}_{\mathrm{D2}}^\mathrm{T}\mathbf{X}_{\mathrm{D2}}\right)^{-1} \end{bmatrix}.$$

Hence, the WLS estimator of Equation C2 can be written as

$$\hat{\boldsymbol{\beta}}_W = \begin{bmatrix} \left(w_2^2\mathbf{X}_{\mathrm{D2}}^\mathrm{T}\mathbf{X}_{\mathrm{D2}}\right)^{-1}\left(w_2^2\mathbf{X}_{\mathrm{D2}}^\mathrm{T}\mathbf{Y}_2\right) \\ \left(w_1^2\mathbf{X}_{\mathrm{D1}}^\mathrm{T}\mathbf{X}_{\mathrm{D1}}\right)^{-1}\left(w_1^2\mathbf{X}_{\mathrm{D1}}^\mathrm{T}\mathbf{Y}_1\right) - \left(w_2^2\mathbf{X}_{\mathrm{D2}}^\mathrm{T}\mathbf{X}_{\mathrm{D2}}\right)^{-1}\left(w_2^2\mathbf{X}_{\mathrm{D2}}^\mathrm{T}\mathbf{Y}_2\right) \end{bmatrix} = \begin{bmatrix} \hat{\boldsymbol{\beta}}_2 \\ \hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_2 \end{bmatrix}, \tag{C5}$$

where $\hat{\boldsymbol{\beta}}_1$ and $\hat{\boldsymbol{\beta}}_2$ are the separate OLS estimators of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ given in Equation A1. Moreover, the estimated variance and covariance matrix is

$$\hat{\mathbf{V}}_\mathrm{W} = \hat{\sigma}_\mathrm{W}^2 \begin{bmatrix} \left(w_2^2\mathbf{X}_{\mathrm{D2}}^\mathrm{T}\mathbf{X}_{\mathrm{D2}}\right)^{-1} & -\left(w_2^2\mathbf{X}_{\mathrm{D2}}^\mathrm{T}\mathbf{X}_{\mathrm{D2}}\right)^{-1} \\ -\left(w_2^2\mathbf{X}_{\mathrm{D2}}^\mathrm{T}\mathbf{X}_{\mathrm{D2}}\right)^{-1} & \left(w_1^2\mathbf{X}_{\mathrm{D1}}^\mathrm{T}\mathbf{X}_{\mathrm{D1}}\right)^{-1} + \left(w_2^2\mathbf{X}_{\mathrm{D2}}^\mathrm{T}\mathbf{X}_{\mathrm{D2}}\right)^{-1} \end{bmatrix}. \tag{C6}$$

Let $\hat{\boldsymbol{\beta}}_\mathrm{W} = [\hat{\beta}_{\mathrm{W}20}, \hat{\beta}_{\mathrm{W}21}, \hat{\delta}_{\mathrm{W}0}, \hat{\delta}_{\mathrm{W}1}]^\mathrm{T}$ with $\hat{\delta}_{\mathrm{W}0} = \hat{\beta}_{\mathrm{W}10} - \hat{\beta}_{\mathrm{W}20}$, $\hat{\delta}_{\mathrm{W}1} = \hat{\beta}_{\mathrm{W}11} - \hat{\beta}_{\mathrm{W}21}$. It is clear from Equation C5 that the WLS estimators of (β_{10}, β_{11}) and (β_{20}, β_{21}) are identical to the separate OLS result of Equation A1 described in Appendix A. More importantly, the WLS estimator $\hat{\delta}_{\mathrm{W}1}$ of $\delta_1 = \beta_{21} - \beta_{11}$ coincides the difference of the two OLS estimators regardless of the relative weights:

$$\hat{\delta}_{\mathrm{W}1} = \hat{\beta}_{11} - \hat{\beta}_{21}. \tag{C7}$$

However, it follows from Equation C6 that the estimated variance of $\hat{\delta}_{\mathrm{W}1}$ is

$$\hat{\mathrm{V}}(\hat{\delta}_{\mathrm{W}1}) = \hat{\sigma}_\mathrm{W}^2\left\{\frac{1}{w_1^2 SSX_1} + \frac{1}{w_2^2 SSX_2}\right\}, \tag{C8}$$

and the explicit formulation of $\hat{\mathrm{V}}(\hat{\delta}_{\mathrm{W}1})$ ultimately depends on the designated weights of w_1 and w_2.

To neutralize the variance heterogeneity, a natural choice of weights (w_1, w_2) is the square root of the inverse of respective unbiased variance estimator (m_1, m_2), where

$$m_i = \left\{\frac{N_i - 2}{SSE_i}\right\}^{1/2} = \left\{\frac{1}{\hat{\sigma}_i^2}\right\}^{1/2}, \; i = 1 \text{ and } 2. \tag{C9}$$

Thus, it can be shown with the weights defined in Equation C9 that SSE_W and $\hat{\sigma}_\mathrm{W}^2$ in Equation C3 are greatly simplified as $SSE_\mathrm{W} = m_1^2 SSE_1 + m_2^2 SSE_2 = N - 4$ and $\hat{\sigma}_\mathrm{W}^2 = 1$, respectively. Moreover, the resulting estimated variance $\hat{\mathrm{V}}_m$ of $\hat{\delta}_{\mathrm{W}1}$ obtained by Equation C8 is

$$\hat{\mathrm{V}}_m = \frac{\hat{\sigma}_1^2}{SSX_1} + \frac{\hat{\sigma}_2^2}{SSX_2}. \tag{C10}$$

Then, more informatively, it can be readily seen from Equations C7 and C10 that the test T_W for the inference of the coefficient parameter of δ_1 given in Equation C4 can be expressed as

$$T_m = \frac{\hat{\beta}_{11} - \hat{\beta}_{21}}{\left(\hat{\sigma}_1^2 / SSX_1 + \hat{\sigma}_2^2 / SSX_2\right)^{1/2}}. \tag{C11}$$

Consequently, the WLS-based partial F test statistic (MWLS) described in Overton (2001), denoted by our notation F_m here, can be expressed as $F_m = T_m^2$, which in turn follows approximately an F distribution with degrees of freedom 1 and *dfw* under the null hypothesis H_0: $\delta_1 = 0$.

Alternatively, the relative weights based on an unbiased estimator for the reciprocal of respective variance can be considered

$$\omega_i = \left\{\frac{N_i - 4}{SSE_i}\right\}^{1/2} = \left\{\frac{N_i - 4}{(N_i - 2)\hat{\sigma}_i^2}\right\}^{1/2}, \; i = 1 \text{ and } 2. \tag{C12}$$

APPENDIX C (Continued)

In this situation, the obtained estimator of δ_1 remains as $\hat{\delta}_{W1} = \hat{\beta}_{11} - \hat{\beta}_{21}$, as given above in Equation C7. Again, it is equivalent to the difference of the two OLS estimators. Furthermore, it can be shown with the relative weights in Equation C12 that $SSE_W = \omega_1^2 SSE_1 + \omega_2^2 SSE_2 = N - 8$, $\hat{\sigma}_w^2 = (N - 8)/(N - 4)$ and the estimated variance $\hat{V}_\omega$ of $\hat{\delta}_{W1}$ is

$$\hat{V}_\omega = \frac{N-8}{N-4}\left\{\frac{(N_1-2)\hat{\sigma}_1^2}{(N_1-4)SSX_1} + \frac{(N_2-2)\hat{\sigma}_2^2}{(N_2-4)SSX_2}\right\}. \tag{C13}$$

Hence, the corresponding test statistic for H_0: $\delta_1 = 0$ is denoted by

$$T_\omega = \frac{\hat{\beta}_{11} - \hat{\beta}_{21}}{\hat{V}_\omega^{1/2}}. \tag{C14}$$

In contrast, the particular results for OLS regression analysis of the models in Equation 2 or C1 can be viewed as a special case of WLS analysis by setting $w_1 = w_2 = 1$. Hence, it can be immediately obtained from Equations C5 and C6 that the OLS estimator $\hat{\boldsymbol{\beta}}_O = \hat{\boldsymbol{\beta}}_W$ and the estimated variance and covariance matrix

$$\hat{\mathbf{V}}_O = \hat{\sigma}_O^2 \begin{bmatrix} \left(\mathbf{X}_{D2}^T\mathbf{X}_{D2}\right)^{-1} & -\left(\mathbf{X}_{D2}^T\mathbf{X}_{D2}\right)^{-1} \\ -\left(\mathbf{X}_{D2}^T\mathbf{X}_{D2}\right)^{-1} & \left(\mathbf{X}_{D1}^T\mathbf{X}_{D1}\right)^{-1} + \left(\mathbf{X}_{D2}^T\mathbf{X}_{D2}\right)^{-1} \end{bmatrix},$$

where $\hat{\sigma}_O^2 = SSE_O/(N - 4)$ and $SSE_O = SSE_1 + SSE_2$. Certainly, $\hat{\mathbf{V}}_O$ and $\hat{\mathbf{V}}_W$ are markedly different. Specifically, the OLS-based test for H_0: $\delta_1 = 0$ is of the form

$$T_O = \frac{\hat{\beta}_{11} - \hat{\beta}_{21}}{\left\{\hat{\sigma}_O^2(1/SSX_1 + 1/SSX_2)\right\}^{1/2}}. \tag{C15}$$

APPENDIX D
The SAS Program for OLS, WLS, and Welch Procedures

```
DATA BRM;INPUT Y X @@;
*REQUIRED USER SPECIFICATIONS PORTION;
*SPECIFY THE NUMBER OF CASES FOR THE DEFINITION OF MODERATOR Z;
IF _N_ < 13 THEN Z=1;ELSE Z=0;XZ=X*Z;
*SPECIFY THE DATA IN TERMS OF PAIRED-VALUES OF Y AND X SEQUENTIALLY;
DATALINES;
 0.8427  -0.6411  0.7367   -0.4710  1.3724   -0.0829  1.0208   -0.9724  1.0173   1.9320   -0.2700
-0.5577  -1.1647  0.5300   -1.5785  -0.7436  -0.5066  -0.0115  -0.3525  0.5926   0.5504   0.1367
1.7815   2.0942   20.6221  5.8494   0.0773   -4.6409  7.2835   4.0269   -4.8243  2.3101   -8.1438
-5.7033  8.1696   1.0393   0.3258   3.3170   -11.5477 -5.2614  -3.7603  -0.3760  11.9974  7.8449
-10.0975 -5.0683  -1.3754  -1.4049  -9.6727  -4.3949  -18.7952 -3.3753  4.8187   -3.8163  17.8028
5.2447   1.5166   -1.7987  -18.1552 0.2492   16.7311  -0.5163  -6.6831  -0.6478  4.1600   -4.0378
-20.5484 -4.2237  -3.4620  -5.1611  5.8139   11.1631  -2.2820  7.1399   -5.3675  1.6372   3.1738
4.3391   2.8675   -3.2008  -0.0571  -2.4078  -5.4507  1.3559   -5.1712  4.8134   17.2321  2.4378
-2.5326  0.3754   -11.2807 -3.6007  -0.4768  2.7956   -9.5239  -5.8884  -5.5169  -2.3669  11.5640
3.7946   4.0643   1.2667   12.5105  1.5345   11.6301  3.3982   -1.6980  -2.0588  -9.9924  -0.7673
4.4596   7.6824   10.2650  3.2056   -7.4692  0.3997   -4.5522  4.5721   -3.7881  -3.0503  10.4150
8.9247   5.1856   4.9535   -0.6495  -4.8053  -4.6085  3.9554   1.7367   -0.4456  6.5870   2.2837
-5.1916  -0.4220  -5.2506  -0.1328  10.1324  0.4992   -5.8459  0.7123   1.3268   3.2873   -3.5859
-3.8313
;
*END OF REQUIRED USER SPECIFICATIONS;
PROC SORT;BY Z;
PROC REG DATA=BRM NOPRINT TABLEOUT OUTEST=W;MODEL Y=X/MSE;BY Z;
DATA W1(KEEP=Z DF MSE BETAH);SET W;IF _TYPE_='PARMS';RENAME
_EDF_=DF _MSE_=MSE X=BETAH;
DATA W2(KEEP=Z STDERR);SET W;IF _TYPE_='STDERR';RENAME X=STDERR;
DATA W3;MERGE W1 W2;WLSW=(DF-2)/(DF*MSE);
*OLS;
PROC REG DATA=BRM;MODEL Y= X Z XZ;
*WLS;
DATA WLS;MERGE BRM W3;BY Z;
PROC REG DATA=WLS;MODEL Y=X Z XZ;WEIGHT WLSW;
```

APPENDIX D (Continued)

```
*WELCH;
DATA WELCH(KEEP=DF1 DF2 STDERR1 STDERR2 BETAH1 BETAH2 WELCH_T
WELCH_DF WELCH_PVALUE);
SET W3;IF Z=1;DF1=DF;STDERR1=STDERR;BETAH1=BETAH;
SET W3;IF Z=0;DF2=DF;STDERR2=STDERR;BETAH2=BETAH;
WELCHVAR=STDERR1**2+STDERR2**2;APS1=(STDERR1**2)/WELCHVAR;
WELCH_T=(BETAH1-BETAH2)/SQRT(WELCHVAR);
WELCH_DF=1/((APS1**2)/DF1+((1-APS1)**2)/DF2);
WELCH_PVALUE=2*(1-PROBT(ABS(WELCH_T),WELCH_DF));
PROC PRINT;
```

APPENDIX E
The R Program for OLS, WLS, and Welch Procedures

```
welch=function ()
{
#REQUIRED USER SPECIFICATIONS PORTION
#SPECIFY THE FIRST GROUP OF DATA IN TERMS OF PAIRED-VALUES OF Y AND X
SEQUENTIALLY
yx1=c(0.8427, -0.6411, 0.7367, -0.4710, 1.3724, -0.0829, 1.0208, -0.9724, 1.0173, 1.9320,
-0.2700, -0.5577, -1.1647, 0.5300, -1.5785, -0.7436, -0.5066, -0.0115, -0.3525, 0.5926,
0.5504, 0.1367, 1.7815, 2.0942)
#SPECIFY THE SECOND GROUP OF DATA IN TERMS OF PAIRED-VALUES OF Y
AND X SEQUENTIALLY
yx2=c(20.6221, 5.8494, 0.0773, -4.6409, 7.2835, 4.0269, -4.8243, 2.3101, -8.1438, -5.7033,
8.1696, 1.0393, 0.3258, 3.3170, -11.5477, -5.2614, -3.7603, -0.3760, 11.9974, 7.8449,
-10.0975, -5.0683, -1.3754, -1.4049, -9.6727, -4.3949, -18.7952, -3.3753, 4.8187, -3.8163,
17.8028, 5.2447, 1.5166, -1.7987, -18.1552, 0.2492, 16.7311, -0.5163, -6.6831, -0.6478,
4.1600, -4.0378, -20.5484, -4.2237, -3.4620, -5.1611, 5.8139, 11.1631, -2.2820, 7.1399,
-5.3675, 1.6372, 3.1738, 4.3391, 2.8675, -3.2008, -0.0571, -2.4078, -5.4507, 1.3559,
-5.1712, 4.8134, 17.2321, 2.4378, -2.5326, 0.3754, -11.2807, -3.6007, -0.4768, 2.7956,
-9.5239, -5.8884, -5.5169, -2.3669, 11.5640, 3.7946, 4.0643, 1.2667, 12.5105, 1.5345,
11.6301, 3.3982, -1.6980, -2.0588, -9.9924, -0.7673, 4.4596, 7.6824, 10.2650, 3.2056,
-7.4692, 0.3997, -4.5522, 4.5721, -3.7881, -3.0503, 10.4150, 8.9247, 5.1856, 4.9535,
-0.6495, -4.8053, -4.6085, 3.9554, 1.7367, -0.4456, 6.5870, 2.2837, -5.1916, -0.4220,
-5.2506, -0.1328, 10.1324, 0.4992, -5.8459, 0.7123, 1.3268, 3.2873, -3.5859, -3.8313)

#END OF REQUIRED USER SPECIFICATION

mxy1=matrix(yx1,length(yx1)/2,2,byrow=TRUE)
mxy2=matrix(yx2,length(yx2)/2,2,byrow=TRUE)
x1=mxy1[,2]
y1=mxy1[,1]
x2=mxy2[,2]
y2=mxy2[,1]
n1=length(x1)
n2=length(x2)
z1=rep(1,n1)
z2=rep(0,n2)
y=c(y1,y2)
x=c(x1,x2)
z=c(z1,z2)
#OLS
summary_ols=summary(lm(y~1+x+z+x:z))
print('OLS')
print(summary_ols$coefficients)
summary1=summary(lm(y1~1+x1))
betah1=summary1$coefficients[2,1]
stderr1=summary1$coefficients[2,2]
df1=summary1$df[2]
mse1=summary1$sigma^2
summary2=summary(lm(y2~1+x2))
betah2=summary2$coefficients[2,1]
```

APPENDIX E (Continued)

```
stderr2=summary2$coefficients[2,2]
df2=summary2$df[2]
mse2=summary2$sigma^2
#WLS
wlsw=c(rep((df1-2)/(df1*mse1),n1),rep((df2-2)/(df2*mse2),n2))
summary_wls=summary(lm(y~1+x+z+x:z,weights=wlsw))
print('WLS')
print(summary_wls$coefficients)
#WELCH
welchvar=stderr1^2+stderr2^2
aps1=(stderr1^2)/welchvar
welch_t=(betah1-betah2)/sqrt(welchvar)
welch_df=1/((aps1^2)/df1+((1-aps1)^2)/df2)
welch_pvalue=2*(1-pt(abs(welch_t),welch_df))
print('Welch Procedure')
print(c('Welch_t','Welch_df','Welch_pvalue'))
print(c(welch_t,welch_df,welch_pvalue))
}
```

(Manuscript received March 27, 2008;
revision accepted for publication August 29, 2008.)