Models of Moderation in Regression

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE
UNIVERSITY OF MINNESOTA

BY

Yi Kang

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Mark L. Davison
Ernest C. Davenport Jr.

July 2024

Abstract

Moderated regression is used when there are continuous variables and categorical variables in the regression model. The purpose of moderated regression is to explain the differences between the levels of the categorical variables. This study introduces a set of regression models with and without moderating effects. To evaluate the effectiveness of those models, a simulation study was used. Type I error rates, power, and identification rates for the generating models are reported. Results show that a small difference between degrees of freedom for two models is related to inflated Type I error rates and sample size 500 is usually associated with adequate power and identification of the generating model. A large difference of *R* values between the focal group and the reference group is related to adequate power.

Table of Contents

Introduction

Regression problems with continuous variables and categorical variables are commonly used in educational studies. When there is interest between the association of a continuous variable and the outcome variable, and a categorical variable may change the strength of the relation between the continuous variable and the outcome variable, the categorical variable is called the moderator and moderated regression is used. For example, Saunders (1956) developed a model with interaction terms including all the effects for all subgroups. The formula for moderated regression presented by Saunders is:

$$y = \bar{y} + \sum_i a_i x_i + \sum_j b_j z_j + \sum_{i,j} c_{ij} x_i z_j$$

In the equation presented above, the z_j s, are the variables that moderate the relation between the predictor x_i s and the criterion variable y. The z_j s are the moderator variables. The products $x_i z_j$ are interaction terms.

Because researchers thought that the model described in Saunder's study (1956) may lead to problems of interpretation (Cook & Weisberg,2004), a few new models were developed. For example, Cook and Weisberg (2004) developed a Partial-One Dimensional Model. That model uses a multiplicative constant to account for the existence of the interaction. Similarly, the Mean Moderated Model (Davison, Davenport, and Kohli, 2017) was also developed. The Linearly Moderated Model, a

combination of the Partial-One Dimensional model and the Mean Moderated Model. was also proposed.

Park (2018) systematically reviewed the models mentioned above including the Partial-One Dimensional Model, the Mean Moderated Model, the Linearly Moderated Model and the model discussed by Saunders. By using a real data set, the author compared pairs of models and fit all models to the data set.

However, since Park's study (2018) was an empirical study, the author listed the results but did not provide a guide to selecting the "optimal" model. To answer the question of "how to select an optimal model", a simulation study is required.

The purpose of this thesis is to use the simulation method to evaluate the most powerful model to detect the interaction when two models are compared and the identification of the generating model when one model is compared with others simultaneously. These models begin with the regression model without moderating effects and end with the Full Moderated Model suggested by Saunders (1956). Intermediate models nested in the Full Moderated Model are also discussed.

In this thesis, models with and without varying intercepts are included. Models from the literature with moderating effects are described: a Partial One- Dimensional model (Cook & Weisberg, 2004) with and without varying intercepts, the Mean Moderated Model with and without intercepts (Davison, Davenport, & Kohli, 2017), and the Full Moderated Model with or without varying intercepts (Saunders, 1956).

Two new models, the Linearly Moderated Model with and without varying intercepts, are introduced.

After reviewing ten models, the models without moderating effects and the models with moderating effects and with varying intercepts are used in the simulation study. Since the Linearly Moderated Model does not exist in any literature except for Park's (2018) dissertation. It will not be included in the simulation study.

As stated previously, Park (2018) did not provided a guide for selecting the "optimal" model when two models were compared or one model was compared with all the other models. This thesis will attempt to answer these questions by simulating results for 5 model :

1. When two models are compared using the *F* statistic, what will be the Type I error rate when the generating model is the less general model of the two fitted models.
2. When two models are compared with the *F* statistic, what will be the power when the generating model is the more general of the two fitted models?
3. If the hypothesis is a model, how well can we identify the generating model when the *F* statistic is used to compare several models?

Review of Literature

In this chapter, real data studies and simulation studies of moderation in regression will be reviewed.

Methodological Review

In this methodological review, there are two models without interaction terms and eight models to detect the existence of interaction terms, and each model will be reviewed systematically. Then, the model selection method will also be reviewed.

The models differ in their regression weights (regression coefficients) in the reference and focal groups. The relationships between the regression weights of the reference group and focal groups differ on four properties. First, the regression weights can differ in their overall mean, here called "level". Second, the regression weights can differ in the size of intervals between regression weights. That is, for any two predictors *v* and *v'*, the difference $\beta_v - \beta_{v'}$ may be different in the reference and focal groups. Third, the signs may be different. That is, β_v may have a different sign in the reference and focal groups. Finally, the rank order of the regression weights may differ in the focal and reference groups, here called "order". These properties will be used to compare the various models in this study. They were also used to develop hypotheses. These properties may help researchers in comparing the empirical regression weights in their studies. These properties may also be useful in deciding which models are most plausible for a given set of data.

Ten Models

In describing the ten models, I will start from the regression model without interaction terms (also called the base model or EWEI). Then, I will examine some existing multiple regression models: the ANCOVA (also called the EWUI) model and

the Full Moderated Model with equal intercepts and unequal intercepts. Finally, I will introduce six new moderated models which have fewer parameters to estimate than in the Full Moderated Model (FMM) but are more complex than the base model.

Table 1 summarizes the ten models.

Table 1

The Equations for the Ten Models Used In this Study

Equation	Name	Acronym
$Y = \sum_v \beta_v X_v + \beta_0 + e$	Equal Weights Equal Intercept Model	EWEI
$Y = \sum_v \beta_v X_v + \sum_{g=1}^{g=G} \beta_g z_g + \beta_0 + e$	Equal Weights Unequal Intercepts Model	EWUI/ANCOVA
$Y = \sum_v \beta_v X_v + \sum_{g=1}^{g=G} k_g \sum_v \beta_v z_g X_v + \beta_0 + e$	Partial One-Dimensional Model with Equal Intercepts	PODEI
$Y = \sum_v \beta_v X_v + \sum_{g=1}^{g=G} \sum_v k_g \beta_v z_g X_v + \sum_{g=1}^{g=G} \beta_{0g} z_g + \beta_0 + e$	Partial One-Dimensional Model with Unequal Intercepts	PODUI
$Y = \sum_v \beta_v X_v + \sum_{g=1}^{g=G} \beta_g^* z_g (\sum_v X_v) + \beta_0 + e$	Mean Moderated Model with Equal Intercept	MMMEI
$Y = \sum_v \beta_v X_v + \sum_{g=1}^{g=G} \beta_g^* z_g (\sum_v X_v) + \sum_{g=1}^{g=G} \beta_{0g} z_g + \beta_0 + e$	Mean Moderated Model with Unequal Intercepts	MMMUI
$Y = \sum_v \beta_v X_v + \sum_{g=1}^{g=G} \sum_v k_g \beta_v z_g X_v + \beta_g^* z_g (\sum_v X_v) + \beta_0 + e$	Linearly Moderated Model with Equal Intercept	LMMEI

Models of Moderation in Regression

Equation	Name	Acronym
$Y = \sum_v \beta_v X_v + \sum_{g=1}^{g=G} k_g \beta_v z_g X_v + \beta_g^* z_g (\sum_v X_v) + \sum_{g=1}^{g=G} \beta_{0g} z_g + \beta_0 + e$	Linearly Moderated Model with Unequal Intercepts	LMMUI
$Y = \sum_v \beta_v X_v + \sum_{g=1}^{g=G} \sum_v \beta_{vg} z_g X_v + \beta_0 + e$	Full Moderated Model with Equal Intercept	FMMEI
$Y = \sum_v \beta_v X_v + \sum_{g=1}^{g=G} \sum_v \beta_{vg} z_g X_v + \sum_{g=1}^{g=G} \beta_{0g} z_g + \beta_0 + e$	Full Moderated Model with Unequal Intercepts	FMMUI

Equal Weights Equal Intercept Model. The base model, Equal Weights Equal Intercept (EWEI), is a multiple regression model, an extension of the simple linear regression model, which can be written as:

$$Y = \sum_v \beta_v X_v + \beta_0 + e \quad (1)$$

$$DF_1 = V,\ DF_2 = N - (V + 1),$$

where β_v is the regression weight for the continuous predictor X_v, β_0 is the intercept, and e is the error term. Given that there are V continuous predictors, there are V+1 parameters to be estimated. The degrees of freedom of the model is V, and the degrees of freedom of residuals is N-(V+1).

In the EWEI model, the regression weights and intercept are assumed to be the same across groups. With respect to the four properties described above -- the level, intervals, signs, and rank order of predictor coefficients -- are the same in the

reference and all focal groups. In other words, this model assumes that all the participants are from the same population. This base model can be compared with the other models, which all take the effect of subgroups into account.

The EWUI Model (ANCOVA Model). The Equal Weights Unequal Intercepts (EWUI) model, more commonly called the analysis of covariance (ANCOVA) model, combines regression analysis and analysis of variance. Generally, ANCOVA is an extension of the ANOVA model, which compares the means of different groups and assumes that the qualitative group variable explains variation in the response variable. The ANCOVA model assumes that, in addition to the qualitative variable in the ANOVA model, there are quantitative variables related to the response variable.

The ANCOVA model for multiple focal groups g (g = 1, …, G) can be written as:

$$Y = \sum_v \beta_v X_v + \sum_{g=1}^{g=G} \beta_g z_g + \beta_0 + e \quad (2)$$

$$DF_1 = V + G,\ DF_2 = N - (G + V + 1),$$

where β_v is the regression weight for the continuous variable X_v in all groups, Z_g is a dummy coded variable with the g^{th} focal group coded as 1 and other focal groups coded as 0, $\sum_{g=1}^{g=G} \beta_g z_g$ is the intercept difference of each focal group from the reference group, and β_0 is the intercept of the reference group. The $\sum_{g=1}^{g=G} \beta_g z_g$ can be interpreted as treatment effects, which indicates the difference of group intercepts.

Given that there are V predictors and G focal groups, there are V+G+1 parameters to estimate. The degrees of freedom of the model is V+G, and the degrees of freedom of residuals is N-(G+V+1).

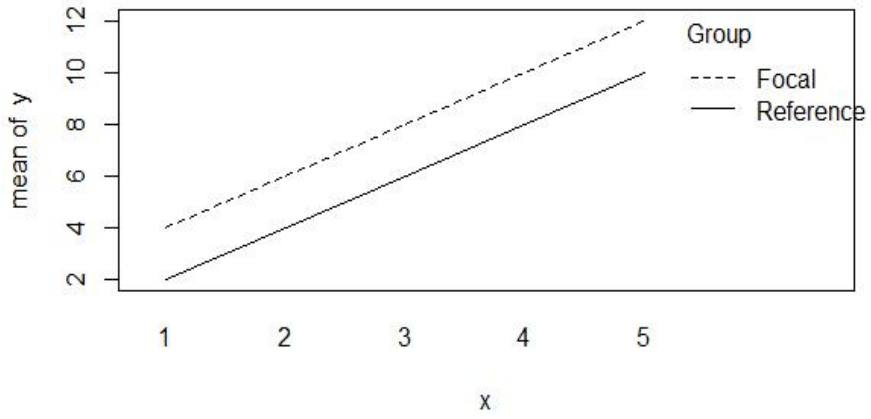
There are several key assumptions for the ANCOVA model: linearity of regression, homogeneity of error variances across groups, independence of error terms, normality of error terms and homogeneity of regression slopes. The last assumption means that there are no interactions between the categorical variable and the covariates. Graphically, there are G+1 parallel lines to represent the regression line for each group, where G is the number of focal groups.

Graphically, there are $G+1$ parallel lines to represent the regression line for each group, where G is the number of focal groups.

Figure 1 shows an example of the ANCOVA model for a single continuous predictor and a single qualitative variable with two levels. These two levels represent a reference and a focal group, which leads to two regression lines with the same slopes but different intercepts. The difference of the intercepts can be interpreted as the difference between two expected values of Y for each possible vector of predictors $\boldsymbol{X}$ in the reference and focal groups.

Figure 1

Example of ANCOVA Model with Two Groups



For the ANCOVA model, the treatment effect, coded as a dummy variable, and the constancy of intercepts and slopes can be tested separately. If there is no evidence to reject the constancy of intercepts null hypothesis, the conclusion is that there is no difference of intercepts between groups. The ANCOVA model becomes the base model.

For the treatment effects, the procedure is to test whether the treatments have any effect, or if all the groups have the same intercept. The null hypothesis and alternative hypotheses for each treatment effect are the same as for ANOVA.

For the constancy of slopes, the ANCOVA model should be compared with the models with interaction terms, which will be discussed in the next section. Like the EWEI model, the ANCOVA model assumes that the regression weights for the *V* predictors have the same level, intervals, signs, and rank order in the reference and all focal groups. As defined at the beginning of the literature review , the “level” refers to

the overall mean. The interval refers to the difference $\beta_v - \beta_{v'}$ in the reference and focal groups. β_v should have the same sign in the reference and focal group. That is both of them are either negative or positive. The order of the regression weights should be the same in the focal and reference group.

Full Moderated Model. Equation (3) describes the Full Moderated Model (FMM) with equal intercepts, which allows slopes to vary across groups, but the intercepts are the same (FMMEI).

$$Y = \sum_v \beta_v X_v + \sum_{g=1}^{g=G} \sum_v \beta_{vg} z_g X_v + \beta_0 + e \quad (3)$$

$$DF_1 = (G+1)V,\ DF_2 = N - ((G+1)(V)+1),$$

where β_v represents the regression weight in the reference group, β_{vg} is the difference between the regression coefficient for variable X_v in focal group g and the reference group, and β_0 is the intercept in the reference group. Given *G* focal groups and *V* predictors, the number of parameters to be estimated is *(G+1)V*+1, the degrees of freedom of the model is *(G+1)V*, and the degrees of freedom of residuals is *N-(G+1)V*-1.

The FMM can also allow the intercepts to vary across groups (FMMUI). FMMUI is the most complex model and can be written as:

$$Y = \sum_v \beta_v X_v + \sum_{g=1}^{g=G} \sum_v \beta_{vg} z_g X_v + \sum_{g=1}^{g=G} \beta_{0g} z_g + \beta_0 + e \quad (4)$$

$$DF_1 = (G+1)V + G,\ DF_2 = N - (G+1)(V+1),$$

where β_{0g} is the difference between the intercept in focal group g and the reference group, g=0 represents the reference group, and the other terms are similar to those in

equation (3). Each subgroup has separate regression lines with different intercepts and different slopes.

With respect to the four properties described above, the FMMUI can account for changes in level, intervals, signs, and rank orders. With one exception described in the next section, it is the only model that can account for a change in the rank order of regression weights (e.g. if the weight vector in the reference group were (1, 2, 3, 4) and that in the focal group was (1, 3, 4, 2)).

Partial One-dimensional Regression model. Cook and Weisberg (2004) developed a model for one grouping variable and several continuous predictors. The model, called the Partial One-Dimensional (POD) Model, is less complex and more easily interpreted than a FMM. If the multiplicative property holds for the different groups, the POD can be used.

The POD model with a common intercept (PODEI) for all groups can be written as:

$$Y = \sum_v \beta_v X_v + \sum_{g=1}^{g=G} k_g \sum_v \beta_v z_g X_v + \beta_0 + e \qquad (5)$$

$$DF_1 = V + G,\ DF_2 = N - (V + G + 1),$$

where β_v represents the regression weight for variable *v* for the reference group, and β_0 is the common intercept for all groups. The intercept and the multiplicative constant are estimated in two steps. The constant k_g, the multiplicative constant described in step 2, is defined for each focal group.

Given there are *V* predictors and *G* focal groups, there are *V*+*G*+1 parameters to

be estimated in the model. The degrees of freedom of the model is $V+G$, and the degrees of freedom of residuals is $N-(V+G+1)$.

A more complex model with different intercepts varying across groups (PODUI) can be written as:

$$Y = \sum_v \beta_v X_v + \sum_{g=1}^{g=G} \sum_v k_g \beta_v z_g X_v + \sum_{g=1}^{g=G} \beta_{0g} z_g + \beta_0 + e \quad (6)$$

$$DF_1 = V + 2G, \; DF_2 = N - (V + 2G + 1).$$

The interpretation of the parameters β_v and k_g is the same as that in equation 5. The equation $\beta_{vg} = k_g \beta_v$ describes the proportionality between the regression weight difference for the reference group and each focal group. The parameter β_{0g} represents the intercept differences between each focal group and the reference group. Given that there are V predictors and G focal groups, there are $V+2G+1$ parameters to estimate in the model. The degrees of freedom of the model is $V+2G$, and the degrees of freedom of residuals is $N-(V+2G+1)$.

Since the constant k_g represents the adjustment of slopes for each focal group, hypothesis tests can be conducted to determine if the multiplicative constant for each focal group,k_g, is significantly different from 0, which is the multiplicative constant of the reference group. One can also do a model comparison with the FMM to ascertain if a proportional constant is sufficient to represent the regression weights of the focal groups.

Cook and Weisberg (2004) stated that the estimation of parameters could be completed using iterative maximum likelihood with two steps in each iteration. In

step 1, the regression weights for all the predictors and the intercept in the reference group are estimated. The predictors in each focal group are multiplied by a constant estimated in step 2 for each group. In step 2, the predictor is computed from the sum of the product of the regression weights estimated in step 1 and the original predictors, while the dependent variable was the same as the one in step 1. The regression weight estimated for each group is based on the multiplicative constant used in step 1. The procedure is repeated until it converges. Step 1 involves estimating β_v for each variable. Step two involves estimating k_g for each group.

Cook and Weisberg (2004) used maximum likelihood to estimate the regression weights and multiplicative constants in both steps. However, since the other eight models are fitted using the ordinary least squares method, for the purpose of model comparisons, the alternating least squares method was selected for parameter estimation in the POD model. Alternating least squares can be used when two sets of parameters, A and B, are estimated. First, A is fixed and B is optimized. Then, B is fixed and A is optimized. This process can be written as the following equations:

$$Y = \sum_v \beta_v (k_g^* X_v) + \sum_{g=1}^{g=G} \beta_{og} z_g + \beta_0 + e \qquad (7.1)$$

In fitting this step of the model, the predictors are $k_g^* X_v$ and z_g. In the first iteration, k_g^* is set to 1.0, and in succeeding iterations it is set to the estimated value from step 2 of the prior iteration.

Equation (7.1) represents the process of estimating the regression weights when

the multiplicative constant is fixed. In Equation (7.1), the term k_g is fixed. It is estimated in the second step and set as 1 for all the groups as the initial value. The term β_v, representing the regression weights of each group, and β_{0g}, the intercept differences between each focal group and the reference group, are estimated.

Equation (7.2) represents the process of estimating the multiplicative constant k_g after fixing the sum of products between the regression weights and values of predictors. In this step, the terms β_{0g} and $\sum_v \beta_v X_v$ are fixed values (the predictors), and the multiplicative constant, k_g, is estimated.

$$Y = \sum_v \beta_v X_v + k_g \sum_v \beta_v X_v + \sum_{g=1}^{g=G} \beta_{0g} z_g + \beta_0 + e \qquad (7.2)$$

The process is iterative and the estimates are obtained by the ordinary least squares method within each step of iteration. Since the ordinary least squares method minimizes the sum of squares of errors after each iteration, the sum of squares of errors decreases and *R* squared increases. If the sum of squares of errors stays unchanged, the process terminates. The alternating least squares method can be applied to the process of estimating regression weights and multiplicative constants for the POD model. The following conditions are the criteria of convergence: 1) regression weights in step 1 are unchanged between two iterations, 2) *R* squared is unchanged between two iterations, and 3) the estimates of multiplicative constants remain unchanged between two iterations.

The multiplicative constants are estimated in step 2 in the algorithm using the ordinary least squares method. After the estimation is completed, the test statistics for

the regression weights in the POD model follow a *t*-distribution with degrees of freedom *N-(V+G+1)*,where *V* represents the number of predictors and *G* represents the number of focal groups.

If there is evidence that the multiplicative constant is nonzero, different slopes are required for different groups. If the multiplicative constant is not different from 0 for every group, the focal group and the reference group have common slopes. In that case, the PODUI model reduces to the ANCOVA model.

With respect to the criteria discussed above, the POD model predicts that, unless k_g= 0, both the mean and the intervals will change in size. If ($1 + k_g$) is positive, the signs of each regression weight will be the same in both the reference and focal group and the rank order of the regression weights will be the same. However, if $(1 + k_{g)}$ is negative, *all* of the signs will change from the reference to the focal group, and the rank order of weights will reverse completely. The POD model cannot account for the situation in which some, but not all predictor coefficients change sign. It also cannot account for the situation in which there are changes in the rank order of predictors, but it is not a complete reversal of the ordering.

Mean Moderated Model . Davison, Davenport, and Kohli (2017) proposed a model that is called the Mean Moderated Model (MMM). If the additive property holds for the different groups, the MMM can be used. Instead of using a multiplicative constant to describe the difference of regression weights between each

focal group and the reference group, the MMM quantifies the difference of regression weights between each focal group and the reference group by an additive constant.

The model of a focal group can be written as:

$$Y = \sum_v (\beta_v + \beta_g^*)X_v + \beta_{0g} + e = \sum_v \beta_v X_v + \sum_v \beta_g^* X_v + \beta_{0g} + e$$
$$= \sum_v \beta_v X_v + \beta_g^* \sum_v X_v + \beta_{0g} + e \ . \quad (8.1)$$

Equation (8.1) shows that the regression weights between groups differ by an additive constant β_g^*, and the interaction includes the total score.

More simply, equation (8.1) can be written as:

$$Y = \sum_v \beta_v X_v + \sum_{g=1}^{g=G} \beta_g^* z_g(\sum_v X_v) + \beta_0 + e \quad (8.2)$$

$$DF_1 = V + G, \ DF_2 = N - (V + G + 1).$$

The first term is the sum of the product of the regression weights and each predictor in the reference group. The second term is the interaction between the dummy variable and the total score for each observation. The third term is the common intercept for all groups. The model is defined as the Mean Moderated Model with equal intercepts (MMMEI).

Given that there are *V* predictors and *G* focal groups, there are *V+G+1* parameters to estimate. Thus, the degrees of freedom of the model is *V+G*, and the degrees of freedom of residuals is *N-V-G-1*.

Equation (9) describes a model with the same property as described in equation (8.2), but the intercepts are allowed to vary by groups: Mean Moderated Model with unequal intercepts (MMMUI). The model can be written as:

$$Y = \sum_v \beta_v X_v + \sum_{g=1}^{g=G} \beta_g^* z_g (\sum_v X_v) + \sum_{g=1}^{g=G} \beta_{0g} z_g + \beta_0 + e \qquad (9)$$

$$DF_1 = V + 2G, \ DF_2 = N - (V + 2G + 1).$$

The third term describes the difference of intercepts between the reference group and each focal group. Other terms are the same as those in equation (8.2). Given that there are *V* predictors and *G* focal groups, there are *V*+*G*+*G*+1 parameters to estimate. The degrees of freedom of the model is *V*+*2G*, and the degrees of freedom of residuals is *N*-(*V*+*2G*+1).

For the two MMMs, the interaction term between the dummy variable and the total can be interpreted as: one unit change in the predictor variable total score is associated with a change in the expected value of *Y* that is b_g^* units larger (smaller) than the expected change in *Y* from a corresponding one unit increase in the reference group. Increasing the total score by one point means increasing one of the predictor variables by one point, and the expected increase will be the same no matter which predictor variable is increased by one point.

The MMM can be compared with the EWEI model. The null hypothesis can be stated as: all the groups have the same regression weight. The alternative hypothesis can be stated as: not all of the groups have the same regression weight, but they differ by only an additive constant. Equivalently, it can be written as:

$H_0 : \beta_{g^*} = 0$ for all focal groups.

$H_A : \beta_{g^*} \neq 0$ for at least one focal group.

If there is no statistical evidence to reject the null hypothesis, then the

MMMUI will reduce to the ANCOVA model. If there is statistical evidence to reject the null hypothesis, the difference of regression weights between the focal group and the reference group is significant and readily interpreted. One can also do a model comparison with the FMM to ascertain if an additive constant is sufficient to represent the regression weights of the focal groups.

The Mean Moderated Model contains a total score interaction as a predictor. Then, the regression weights of the total score interaction, representing the additive constant, is estimated.

A researcher may be interested in the question of whether the moderator variable has the same effect on every predictor: i.e. is the difference β_v - β_{vg} the same for every variable X_v in group g? The MMM models incorporate the assumption that the difference β_v - β_{vg} is the same for every variable X_v in group g, and therefore it is useful in addressing the hypothesis that the moderator variable has the same effect on every predictor in the focal group.

The Mean Moderated Model can be used to study patterns of predictor scores (Davison & Davenport, 2002; Davison et al., 2022; Wiernik et al., 2021;). Conceptually, a pattern is the configuration of a person's highest and lowest scores. In this conceptualization, Person A with SAT scores Verbal = 400, Quantitative = 600 has a different pattern than does Person B with Verbal = 600, Quantitative = 400. Both have the same total score, 1000, but they have different within person patterns. Davison et al. (2002, 2022) developed a regression method for determining whether

there is a criterion-related pattern, a particular within person configuration of highest and lowest scores associated with high scores on a criterion. They have shown that the MMM can be used to investigate whether the predictor pattern associated with high scores on the criterion is moderated by group membership; e.g. is it the same pattern for males and females? More specifically, if the MMM holds, the criterion-related pattern (if there is one) is the same pattern in both the reference and focal group.

Davison et al. (2023) were interested in whether total math credits in high school was a sufficient statistic for predicting mathematics achievement; or did the content of that coursework have an influence on achievement over and above the effect of total math credits earned? Davison et al. (2023) found a criterion-related pattern of coursework content associated with high math achievement. They then used the Mean Moderated Model to assess whether the criterion-related pattern remained the same across ethnic groups.

In terms of the four properties of how regression weights differ between the reference and focal groups, the MMM models imply that the regression weights for the reference and focal groups differ only in their level. There also may be changes of sign for some or all predictor coefficients. The MMM model cannot account for changes in the size of intervals, and it cannot account for changes in the rank order of the coefficients. If the variance of the predictor coefficients or the rank order changes from reference to focal group, this disconfirms the MMM model.

Linearly Moderated Model. The Linearly Moderated Model (LMM) is a combination of the POD model and the MMM. The regression weights of different groups differ by an additive constant and a multiplicative constant simultaneously. In other words, the difference of regression weights can be written as a linear combination of two constants $\because \beta_{vg} = k_g\beta_v + \beta_g^*$.

Similar to the POD model, the estimation of the multiplicative constants in the LMM depends on a two-stage algorithm. In step 1, the regression weight for all of the predictors and the total score of each observation is estimated. In step 2, the multiplicative constant is estimated for each group. The process is repeated until it converges.

The LMM with a common intercept (LMMEI) for each group can be written as：

$$Y = \sum_v \beta_v X_v + \sum_{g=1}^{g=G} \sum_v k_g \beta_v z_g X_v + \beta_g^* z_g (\sum_v X_v) + \beta_0 + e \qquad (10)$$

$$DF_1 = V + 2G,\ DF_2 = N - (V + 2G + 1).$$

. The model with unequal intercepts (LMMUI) for each group can be written as：

$$\mathrm{Y} = \sum_{\mathrm{v}} \beta_{\mathrm{v}} \mathrm{X_v} + \sum_{\mathrm{g}=1}^{\mathrm{g}=\mathrm{G}} \mathrm{k_g} \beta_{\mathrm{v}} \mathrm{z_g X_v} + \beta_{\mathrm{g}}^* \mathrm{z_g} (\sum_{\mathrm{v}} \mathrm{X_v}) + \sum_{\mathrm{g}=1}^{\mathrm{g}=\mathrm{G}} \beta_{0\mathrm{g}} \mathrm{z_g} + \beta_0 + \mathrm{e} \quad (11)$$

$$DF_1 = V + 3G,\ DF_2 = N - (V + 3G + 1).$$

The POD model and the MMM model are special cases of the LMM model and the FMM. The base model, the ANCOVA model, the MMM and the POD model are all hierarchically embedded in the LMMUI and the FMMUI. If the multiplicative constant is 0, the LMM becomes the MMM. The difference of slopes between the focal group and the reference group is described by an additive constant. If the

additive constant is 0, the LMM becomes the POD model. If both constants are 0, the LMM becomes the ANCOVA model.

With respect to the four properties above, the LMM model can account for changes in level, interval size, and sign between the reference and focal groups. It cannot account for changes in the rank order except in the special case where $(1 + k_g)$ is negative, in which case, the ordering completely reverses from the reference to the focal group.

Model Selection Method

In the previous section, the 10 models were discussed. In this section, the statistics for model comparison and selection will be discussed. Then, I will discuss the method of model comparison as applied to the models with hierarchical structure.

Coefficient of determination. In regression analysis, the coefficient of determination, *R* squared, measures the effect of predictors in accounting for variation in the criterion variable. More precisely, the coefficient of determination can be interpreted as the proportion of the total variation of the response variable explained by the regression model.

The coefficient of determination can be expressed as:

$$R^2 = \frac{SSR}{SST},$$

where *SST* is the sum of squares of the total, and *SSR* is the sum of squares of the regression model.

The coefficient of determination represents the proportion of variance explained by the regression model. The larger the value of the coefficient of determination, the more variation is explained by the predictors in the model.

Moreover, for hierarchically embedded models, the model with more predictors in the regression model will always have a higher value of the coefficient of determination. For the ten models in this study, the model with more interaction terms should have a higher coefficient of determination. Among the 10 models, the EWEI model should have the lowest coefficient of determination, and the FMMUI should have the highest coefficient of determination. For each pair of models, for example, the MMM with and without an intercept varying across groups, the model with different intercepts for focal groups should have a higher coefficient of determination.

A hypothesis test is required to test whether models differing in number of parameters account for equal amounts of variance in the population.

Adjusted *R* squared. An extension of the coefficient of determination is the adjusted *R* squared. It is defined as $1-\frac{(1-R^2)(N-1)}{N-p-1}$, where R^2 is the sample *R* squared, *p* is the number of predictors, and *N* is the total sample size. The value of the coefficient of determination will increase automatically when more predictors are added. The adjusted *R* squared modifies this problem by taking the number of predictors into account. The adjusted *R* squared increases with the increase of predictors if the increase in prediction offsets the loss in degrees of freedom of

residuals.

Hypothesis testing for interaction terms. The general approach for testing the significance of some predictors in the regression model is the *F* test for model comparison.

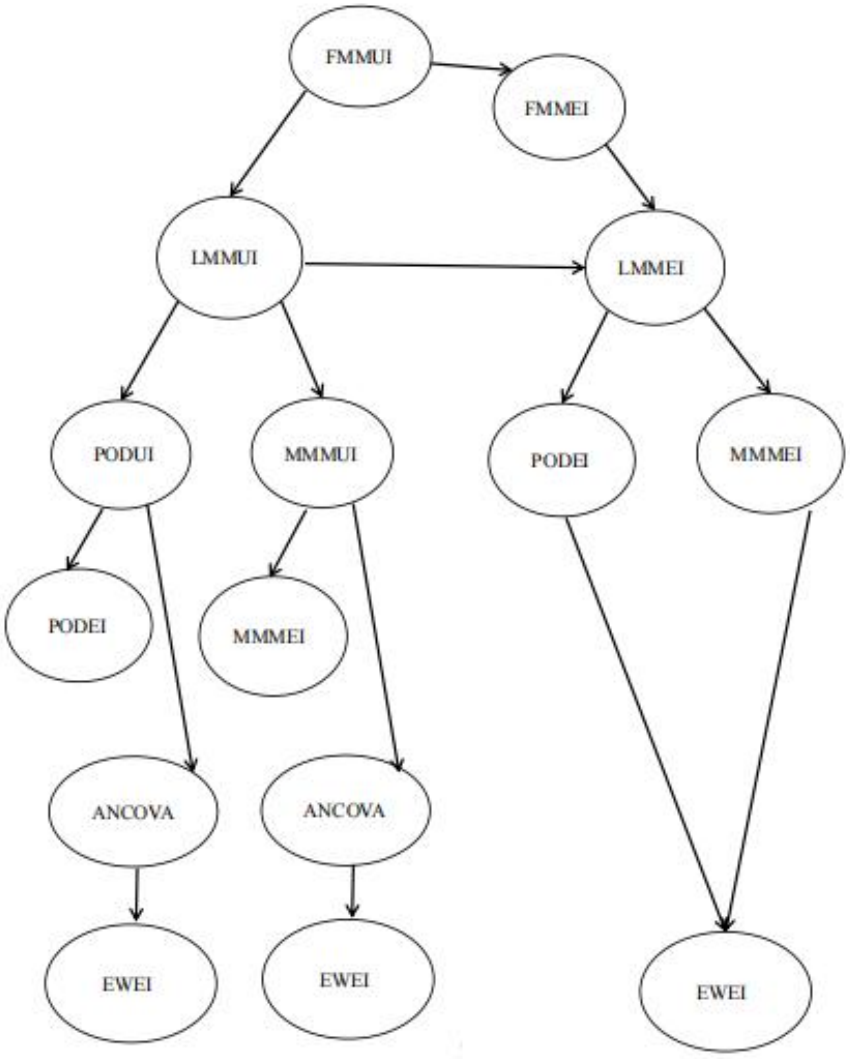
Suppose there are two regression models. If both contain the same predictors but one has at least one additional predictor, the two models are nested. The model with additional terms is defined as the full model or unrestricted model. The model with fewer terms is defined as the reduced model or restricted model.

To test whether one or more predictors in the full model account for more variability of the response variable in the population, the *F* statistic will be computed based on the difference of the sum of squares of errors between the full model and the reduced model as shown below in equation 12. If a relatively small difference in the sum of squared errors between the two models indicates that the additional predictor(s) do not reduce the error variation of the response variable appreciably, then, the null hypothesis of the reduced model cannot be rejected.

This approach can be applied to the ten models to test whether the interaction terms should be retained. Since the *F* test requires a nested structure between two models, not all pairs of models can be compared using the *F* statistic. For example, some models with the equal intercept across groups may not be compared with the models with unequal intercepts across groups, though the former models may have more terms.

Figure 2

The Nested Structure of Ten Models



Among the ten models, the EWEI model is nested in the other nine models, and the other nine models are nested in the FMM with unequal intercepts across groups. The test statistic can be written as:

$$F = \frac{SSE(R) - SSE(F)}{df_E(R) - df_E(F)} / \frac{SSE(F)}{df_E(F)} \quad . \qquad (12)$$

The term *SSE(R)* represents the sum of squares of the reduced model, and the term

SSE(F) represents the sum of squares of the full model. The term $df_E(R)$ represents the degrees of freedom of the residual of the reduced model, and the term $df_E(F)$ represents the degrees of freedom of the residuals of the full model. The test statistic follows an *F* distribution with degrees of freedom $df_E(R)-df_E(F)$ and $df_E(F)$.

Selected Real Data Studies and Simulation Studies

In the previous section, the literature for the ten models to detect interaction was reviewed. In this section, the applications of these models to real and simulated data will be reviewed.

Selected Real Data Studies

Most of the real data studies apply the Full Moderated Model to real data. For example, Saunders (1956) listed some examples using the FMM. The first example involved a data set with 153 observations. The author used a variable "Engineering Score" to predict "Grade Average" by using the regression method. The estimated correlation was 0.102 and statistically significant. Another variable, "the accountant score," measured "compulsiveness." The author split the total group into the "compulsive" group and "non-compulsive" group by using the "accountant score." For the subgroup of "compulsive," the correlation between the variable "Engineering Score" and "Grade Average" was -0.1. Within the subgroup of "non-compulsive," the correlation between the variable "Engineering Score" and "Grade Average" was 0.25. The author used a moderated regression to get a higher correlation 0.153 with two predictors, "the Engineering score" and "the Accountant Score." The moderated

regression equation can be written as:

$$\text{Grade Average} = \beta_0 + \beta_1 * \text{Engineer Score} + \beta_2 * \text{Accountant Score} + \beta_3 * \text{Engineer Score} * \text{Accountant Score} \quad (13)$$

Only a few studies discussed the application of more recently published models such as the POD Model or the Linearly Moderated Model by using real data. Cook and Weisberg (2004) developed the POD model and discussed the reason for using the POD model. Some problems have more than one categorical variable and several covariates making interpretation complex. Though Cook and Weisberg (2004) mentioned the purpose for using the POD model was simplicity and ease of interpretation, the POD model should be used only when the regression weights for different groups are proportional, and the FMM model does not fit significantly better than the POD model.

As described in the previous section, the POD model was fitted by using a two-step iterative algorithm. In step 1, the goal was to estimate regression weights given that the multiplicative constant was supposed to be known. In step 2, the goal was to estimate the multiplicative constant given that the regression weights were known. To compare the POD model with a non-POD model, the likelihood ratio test was used. The data set used in Cook and Weisberg's study was the Australian athlete data with 202 observations. The response variable was the athlete's lean body mass, the covariate was the weight and the categorical variable was the gender. To explore the question of whether the relation between weight and lean body mass was mediated

by the variable gender, the PODUI model was fit. The POD model, set as the full model, was compared to three restricted models: the EWEI model, the ANCOVA model and the common intercepts model. The small *p*-values of likelihood ratio tests showed the existence of group differences.

Park (2018) used real data to compare the performance of the ten models. The purpose of Park's study was to use a set of regression models to study group variation. That study assumed that the relationships between predictors and the criterion differed by groups.

In Park's study, The High School Longitudinal Study of 2009 (HSLS:09) data was used to explore the effects of course-taking on achievement in mathematics for different ethnicity groups. The categorical variable used in the analysis was "ethnicity" with three different subgroups. The subgroup "white" was used as the reference group. The thirteen covariates were the student's grades in thirteen mathematics course areas. The dependent variable was the mathematics achievement score.

To answer the research question about the relation between group differences and the mathematics achievement score, Park began with the EWEI model, then, built the models with more predictors such as the POD Models, the Mean Moderated Models, and the Linearly Moderated Models. Finally, the Full Moderated Models were fitted. All of the ten models were fitted.

To evaluate the effectiveness of each model, the model comparison method and model fit measures, including *R* squared, adjusted *R* squared, *AIC* and *BIC*, were used.

Models were selected via different statistics. Models with the largest values of *R* squared and adjusted *R* squared were defined as the "optimal" ones by Park. Models with the smallest AIC and BIC values were also defined as "optimal" ones by Park. By using different statistics for each analysis, the author might select different models to explain the variation of the mathematics achievement score by ethnicity groups.

Two families of analyses were conducted. The first set of analyses used the 13 coursework areas and ethnicity as independent variables. As stated previously, the author began with the EWEI model which assumed that no group differences existed in the relationship between the independent variable, sequences of course-taking, and the dependent variable, mathematics achievement score. In the regression analysis, no categorical variable was used. Then, the author fitted other models with moderating effects.

In the ANCOVA model, regression coefficients for all three subgroups (Asian, Black/African American and Hispanic) were significantly different from those of Whites. In using the PODEI model, the multiplicative constants for the Black/African American group and the Hispanic group were statistically significant. In using the PODUI model, no multiplicative constants were statistically significant. In using the MMMEI, the additive constants for the Black/African American group and the Hispanic group were statistically significant. In using the MMMUI, the additive constant for the Asian group was statistically significant. In using the LMMEI, additive constants were statistically significant for the Black/African American group

and the Hispanic group. In using the LMMUI, neither additive nor multiplicative constants were statistically significant. In using the FMMEI, the following regression coefficients for interaction terms were significant: “Black/African American group and Geometry” and the “Hispanic group and Geometry”. Further, in using the FMMUI, no regression coefficients for interaction terms were statistically significant. The result of model comparisons indicated that the LMMUI was the best model. The values of adjusted *R* squared also indicated that the LMMUI was the best model. However, the values of AIC and BIC indicated that ANCOVA was the best model. The author presented the result of model comparisons and statistics, but did not provide a general standard to select the “optimal” model. I will revisit this result and provide a summary table in the result part.

The second set of analyses used the 13 coursework areas and ethnicity as independent variables and then added the variable “prior achievement” as a covariate.

The author began with the EWEI model, assuming that there are no ethnic differences in the relationship between prior achievement, sequences of course-taking, and mathematics achievement.

In using the ANCOVA model, the author found evidence that the groups’ mean differences should be considered, since the F-statistic for the model comparison to compare the EWEI model and the ANCOVA model was significant. Then, the author fit the PODEI Model, the PODUI model, the MMMEI, the MMMUI, the LMMEI, the LMMUI, the FMMEI, and the FMMUI.

After fitting the ten models, the author provided evidence for selecting the "optimal" model. The results of model comparisons indicated that the ANCOVA model and the LMMEI were the best models with the consideration of the nested structure. Among the models with unequal intercepts, the author would select the ANCOVA model. Among the models with equal intercepts, the author would select the LMMEI.

Since the ANCOVA model was not nested in the LMMEI, these two models could not be compared by using the *F*-test. The values of adjusted *R* squared indicated that the LMMEI, which had a larger value, was the better model. The values of AIC and BIC indicated that the ANCOVA model, which had lower values of those statistics, was the better model.

As stated above, for the second set of analyses, I will also revisit the result and provide a table in the result part.

Though Park provided an example of applying the ten regression models to the real data, there were some limitations. First, Park did not point out clearly what was the "optimal" model. Second, Park did not explain whether those findings could be applied to other datasets.

Davison, Davenport, & Jia (2022) discussed whether there was a pattern of scores related to the criterion variable, college GPA. The data set was from the CRAN R library (gvlma). The purpose of this study was to set the MMMEI as the null, and the pattern model (such as the FMMEI which contains both the level and pattern) as

the alternative. The covariates were the SAT verbal score and SAT math score. The moderator was gender and the response variable was the first year GPA. As previously, if the MMMEI fit, there was no difference in the pattern of scores related to the criterion variable. The Level Moderated Model, same as the MMMEI, was fitted. The Full Moderated Model was also fitted. The results of the *F* test for model comparison provided evidence that the Full Moderated Model fitted better. Thus, a model containing both the level effect and pattern effect should be used.

Table 2

A Summary Table of "Real Data" Studies

Study	Short Description	Model(s)	Limitations
Saunders (1956)	Used a data set with 153 observations to explore if the magnitude of "Compulsive" impacted the relation between the "Engineer Score" and the outcome variable "Average Grade."	FMM	The author did not provide estimations. It was not compared with other models.
Cook and Weisberg (2004)	Used a data set with 202 observations to explore if the variable "Sex" impacted the relation between the "Weight" and the outcome variable "Lean Body Mass."	POD	The author used the likelihood ratio test to compare the performance of POD with other models. However, the least squared method might be a simpler method .

Study	Short Description	Model(s)	Limitations
Park (2018)	Used a data set with 21389 observations to explore if the variable "Ethnicity" impacted the relation between "thirteen sequences of math courses-taking" and the outcome variable "mathematics achievement" with or without another covariate "prior achievement." Tried to find the optimal model.	EWEI EWUI PODEI PODUI MMMEI MMMUI LMMEI LMMUI FMMEI FMMUI	The author did not provide a generalized standard to select the "optimal"model.
Davison, Davenport, & Jia (2022)	Used a data set provided in R to compare the performance of the "Level Moderated Model" to the "Pattern Moderated Model."	MMMEI FMMEI	Focused on the application of the MMM and the FMM. The POD was not involved.

Review of Previous Simulation Studies

Studies comparing two or more of the ten models to analyze real data were reviewed in the last section. To evaluate the effectiveness of a regression model with a moderation effect, Monte Carlo (MC) methods have been used among the previous

studies.

In some studies, the author compared different methods with a regression method to detect the moderating effect. The research question was: among several methods including the moderated regression method (MMR), which was the most powerful method to detect the existence of the moderating effect?

The regression model used in the studies was the FMM. The equation could be written as:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + e \quad (14)$$

In equation 14, the variable X_1 represents a continuous variable, the variable X_2 represents a categorical variable with two levels, and the product of X_1 and X_2 represents the moderating effect.

For example, the performances of the subgroup-based correlation coefficients between the response variable and the independent variable were compared with the MMR (Stone-Romero & Anderson, 1994). The independent factors used in this study were: different sample sizes, different effect sizes, different reliability levels and different numbers of subgroups. The conclusion was: as compared to the subgroup-correlation coefficients, the moderated multiple regression was more powerful in detecting the moderating effect.

Some studies focus on the on the MMR itself. The research question was: what were the factors impacting the performance of the MMR to detect the interaction?

Most of the studies discussed the factors impacting the power or Type II error

rate. For example, in Stone-Romero, Alliger and Aguinis's (1994) study, the sample size was defined as the total number of observations in the regression analysis with five different sample sizes 30, 60, 90, 180 and 300. The correlation between the dependent variable and independent variable was computed independently for the two groups. In each group, the correlations were set at four values: 0.2, 0.4, 0.6 and 0.8. Both groups used the four values. Thus, there were 16 pairs of correlations. The values of the proportion of people in group 1 was set as 0.1, 0.3 and 0.5. The conclusion was: the larger sample size, the larger difference between group correlation coefficients, and the equal sample sizes in each group were related to higher power.

Aguinis and Stone-Romero (1997) included the total sample size, range restriction (computed by the ratio of the range of the scores in the sample to the range of the scores in the population), proportion of cases in the reference group, the magnitude of the interaction, and the inter-correlation between predictors. The authors set the inter-correlation between two predictors as 0.20, 0.40 and 0.80, and they labeled them as low, medium and high degree of multicollinearity, respectively. They assigned the correlations between the moderator variable Z and the predictor variable X in the reference group, the focal group and the whole group. The conclusion was: under the condition that all the values of inter-correlation between two predictors were set as 0.8, the power was about 0.5 in detecting the interaction for the MMR. That indicated that the higher inter-correlation between predictors might decrease the

power. As stated before, for the previous studies, the total sample sizes, the different sample sizes in different groups, and the correlation between predictors were usually set as the independent factors in previous studies if the MMR was the only method to detect the interaction.

While in most studies, the FMMEI or FMMUI was the only regression model compared with other methods to detect the interactions, Kang (2022) used a simulation study to compare the performances of the EWEI model, the PODEI Model, the MMMEI, and the FMMEI. In that study, the independent factors were different sample sizes and different *R* squared values for generating models. The outcome variables were the Type I error rate and the power. For the PODEI, the null hypothesis is: $H_0 : k = 0$ for all focal groups, and the alternative hypothesis is: $H_A : k \neq 0$ for at least one focal group. For the MMMEI, the null hypothesis is: $H_0 : \beta_g^* = 0$ for all focal groups, and the alternative hypothesis is: $H_A : \beta_g^* \neq 0$ for at least one focal group. For the FMMEI, the null hypothesis is: $H_0 : \beta_{vg} = 0$ for all focal groups, and the alternative hypothesis is: $H_A : \beta_{vg} \neq 0$ for at least one focal group. The result showed that if the generating model was the EWEI, the Type I error rate was inflated if the fitted models were the PODEI and FMMEI. When the generating model was the PODEI, MMMEI or FMMEI, the power decreased to less than 90% under most conditions. When the generating model was the FMMEI, the Type I error rate was inflated to greater than 5%. Though that study provides an example of using simulation study for the "new" models, the major limitation was that the nested

structure of the ten models was not considered.

Table 3 summarizes the results of previous simulation studies.

Table 3

The Previous Simulation Studies

Studies	Purpose	Independent Factors	Conclusion	Limitations
Stone-Romero and Anderson (1994)	To compare the performance of MMR with other methods to detect the interaction.	Different sample sizes, different effect sizes, different reliability levels and different numbers of subgroups.	MMR was the most powerful method.	Some other unusual methods were used. Those methods are not related to the regression models.
Romero, Alliger and Aguinis's (1994) Aguinis and Stone-Romero (1997)	To detect the factors impacting the power of the MMR.	Total sample size, the range restriction, the sample size in each group, the magnitude of the interaction and the inter-correlation between predictors.	The larger sample size, the larger difference between group correlation coefficients, and the equal sample sizes in each group were related to higher power.	FMMUI was the only model included.

Studies	Purpose	Independent Factors	Conclusion	Limitations
Kang (2022)	To compare the performance of the EWEI model, the PODEI model, the MMMEI model and the FMMEI model.	Different sample sizes and different *R* squared for the generating models.	When the generating model was the EWEI, the Type I error rate was inflated if the fitted models were the PODEI and FMMEI. When the generating model was the PODEI, MMMEI and FMMEI, under most conditions, the power decreased compared to 90%. When the generating model was the FMMEI, the Type I error rate was inflated compared to 5%.	The nested structure was not included.

Methodology

In the previous section, I discussed the ten models and examples of simulation studies. Among the ten models, the Linearly Moderated Model only appeared in Park's dissertation. It was not discussed in other literature. I will not use it in my analysis.

In this part, the methodology of this study will be discussed. I will discuss the method of data generation, hypothesis tests, the measures, and other practical issues. Two decision scenarios will be evaluated below. In the first, the researcher is comparing two models, one hierarchically nested in the other, using the *F* statistic. One of the two models is the generating model. The researcher will retain the simpler model (the null model) if the F is not significant at the 0.05 level; otherwise the researcher will reject the null model in favor of the more general, alternative model. Tables 7 – 21 report type I errors for situations in which the generating model is the null model and power for the situation in which the generating model is the alternative. Table 22 reports results related to the second approach, a less familiar strategy in which a given model is the generating model, and it is compared to all of the other models to which it is hierarchically related. The given model is supported if it satisfies two conditions. First, in every comparison in which the given model is the null model, the null hypothesis is retained. Second, in every comparison in which the given model is the alternative, the null hypothesis is rejected. For instance, in this second strategy, the EWEI is compared to four other models: the EWUI, the PODUI, the MMMUI, and the FMMUI. The EWEI is the null model in all four comparisons, so in order for the EWEI to be supported in a replication, the null hypothesis must be retained in all four comparisons. As a second example, if the PODUI is the given model, it is hierarchically related to and compared with three other models: the EWEI, the EWUI, the FMMUI. It is the alternative hypothesis for three comparisons (EWUI

and EWEI), and it is the null hypothesis for one of these comparisons (FMMUI). For the PODUI to be supported by the data in a given replication, the null hypothesis must be rejected in two of the comparisons (EWUI and EWEI), and it must be retained in the third (FMMUI). In this second scenario, the question is whether the generating model will be supported by the data when it is compared to all other models in which it is hierarchically nested.

Data

In this simulation study, there were six sub-studies for each of the two decision scenarios corresponding to the six generating models. Each sub-study had a different generating model: EWEI, ANCOVA, PODUI, MMMUI, or FMMUI with two variations of the FMMUI model described below. Within each sub-study, the independent variables were (a) five fitted models, (b) three sample sizes for each group (N=100, 500 and 1000), and (c) three *R* squared values (0.2, 0.5 and 0.8). This resulted in 3*3*5*6=270 conditions.

The results in a pilot study showed that under the condition that the sample size was 100, the Type I error rate might be inflated and the power would be lower. Thus, in this study, I did not use a sample size lower than 100. Samples sizes of 500 and 1000 were selected to check if those provide adequate power. To make the study manageable, I did not vary the ratio of the reference group and the focal group sample sizes. For each condition, a total of 1000 replications were generated. The method of generating each model is described.

The general approach

The three questions in the study were: (1) In scenario 1, if two models are compared using the *F* statistic, what will be the Type I error rate when the generating model is the less general model of the two fitted models? (2) In scenario 1, if two models are compared with the *F* statistic, what will be the power when the generating model is the more general of the two fitted models? (3) In scenario 2, if we compare the generating model to all of the other models in which it is hierarchically related, how often will the generating model be identified?

The simulation focused on generating models with equal or unequal intercepts and regression coefficients between the reference group and the focal group. The PODUI model and the MMMUI model could not be generated in a direct way. Thus, I used the principle $R^2 = \boldsymbol{r}'_{xy} * \boldsymbol{R}^{-1}_{xx} * \boldsymbol{r}_{xy}$ and $\boldsymbol{\beta} = \boldsymbol{R}^{-1}_{xx'} * \boldsymbol{r}_{xy}$ to get the different regression coefficients for the reference group and the focal group for the PODUI model and the MMMUI.

To give the study better focus, I studied the effects of varying regression weights across groups by varying the vector $\boldsymbol{r}_{xy}$ representing the correlations between *X* and *Y* holding the predictor inter-correlation matrix $\boldsymbol{R}_{xx}$ constant.

For the models with equal regression coefficients for the reference group and the focal group, the $\boldsymbol{r}_{xy}$ values were the same for the reference group and the focal group. For the models with unequal regression coefficients for the reference group and the focal group, the $\boldsymbol{r}_{xy}$ values differed for the reference group and the focal group.

For the data set, the six models mentioned above were generated. For each model, I used four continuous predictors and one categorical predictor to represent two groups. For continuous predictors, I set the correlation equal to 0.3 between all predictors with one another . The coefficient of determination of the generating model was varied: 0.2, 0.5 and 0.8. The *R* squared (0.2, 0.5, 0.8) included a range that would cover most, if not all studies containing four predictors.The corresponding correlation matrices were computed in different ways for different generating models.

Table 4 shows the assumptions for model generation.

Table 4

The Assumptions for the Model Generation

Generating Model	Inter-correlation Matrix $\boldsymbol{R}_{xx}$	$\boldsymbol{r}_{xy}$ Values for R Squared=0.2	R Squared Values	The Values of $\boldsymbol{r}_{xy}$ Used for Computing the Regression Coefficients
EWEI	0.3 for all predictors	(0.3,0.3,0.5,0.5)	0.2 or 0.5 or 0.8	Same for the reference group and the focal group.
EWUI	0.3 for all predictors	(0.3,0.3,0.5,0.5)	0.2 or 0.5 or 0.8	Same for the reference group and the focal group.
PODUI	0.3 for all predictors	(0.3,0.3,0.5,0.5)	0.2 or 0.5 or 0.8	The values of $\boldsymbol{r}_{xy}$ are varied by a multiplicative constant for the reference group and the focal group.
MMMUI	0.3 for all predictors	(0.3,0.3,0.5,0.5)	0.2 or 0.5 or 0.8	The values of $\boldsymbol{r}_{xy}$ are varied by an additive constant for the reference group and the focal group.
FMMUI with Different Rank Orders	0.3 for all predictors	(0.3,0.3,0.5,0.5)	0.2 or 0.5 or 0.8	The $\boldsymbol{r}_{xy}$ of the focal group is permuted as (0.3, 0.3, 0.5, 0.5).

Generating Model	Inter-correlation Matrix $\boldsymbol{R}_{xx}$	$\boldsymbol{r}_{xy}$ Values for R Squared=0.2	R Squared Values	The Values of $\boldsymbol{r}_{xy}$ Used for Computing the Regression Coefficients
FMMUI with the Same Rank Orders	0.3 for all predictors	(0.3,0.3,0.5,0.5)	0.2 or 0.5 or 0.8	The $\boldsymbol{r}_{xy}$ of the focal group is increasing and set as (0.3, 0.4, 0.7, 0.75).

Table 5 summarizes the parameters used for model generation. In the next part, I will describe how the regression coefficients in Table 5 were computed from the values in Table 4.

Table 5

The R Squared Values and the Constant Used for Computing $\boldsymbol{r}_{xy}$

Generating Model	*R* Squared for the Reference Group	*R* Squared for the Focal Group	Constant for Computing the Values of $\boldsymbol{r}_{xy}$ for the Reference Group	Constant for Computing the Values of $\boldsymbol{r}_{xy}$ for the Focal Group
EWEI	0.2	0.2	1	1
	0.5	0.5	1.58	1.58
	0.8	0.8	2.01	2.01
EWUI	0.2	0.2	1	1
	0.5	0.5	1.58	1.58
	0.8	0.8	2.01	2.01
PODUI	0.2	0.5	1	1.58
	0.5	0.8	1.58	2.01
	0.8	0.2	2.01	1
MMMUI	0.2	0.5	0	0.188
	0.5	0.8	0.188	0.320
	0.8	0.2	0.320	0
FMMUI with Different Rank Orders	0.2	0.5	1	1.58
	0.5	0.8	1.58	2.01
	0.8	0.2	2.01	1
FMMUI with the Same Rank Order	0.2	0.5	1	1.58
	0.5	0.8	1.58	2.01
	0.8	0.2	2.01	1

General Principles for Constructing the Regression Weights for the EWEI Model and the POD Model

For the EWEI model and the POD model, I used a similar principle. *R* squared can be written as: $R^2 = \boldsymbol{r}_{xy}^{'} * \boldsymbol{R}_{xx}^{-1} * \boldsymbol{r}_{xy}$ and the regression coefficient can be written as $\boldsymbol{\beta} = \boldsymbol{R}_{xx'}^{-1} * \boldsymbol{r}_{xy}$, where $\boldsymbol{r}_{xy}$ is a somewhat arbitrary initial value of the correlations between the predictors and the response variable from which all other values of $\mathbf{r}_{xy}$ are

computed in the manipulation of regression weights. The $\mathbf{r}_{xy}$ for each group was computed by solving for the value of *k* in the equation $R^2 = \sqrt{k} * \boldsymbol{r}'_{xy} * \boldsymbol{R}^{-1}_{xx'} * \boldsymbol{r}_{xy} * \sqrt{k}$ that yields the specified coefficient of determination for that condition (.2, .5, or .8). I fixed R^2 to one of the values above, then I set all of the values of $\boldsymbol{R}_{xx}$ to 1.0 on the diagonal and 0.3 on the off-diagonal.

Since the regression coefficients can be written as $\boldsymbol{\beta} = \boldsymbol{R}^{-1}_{xx'} * \boldsymbol{r}_{xy}$ for the EWEI and the POD model, I will have different ways to generate these models.

EWEI Generating Model. When the coefficient of determination is 0.2, the principle mentioned above was applied. The $\boldsymbol{R}_{xx}$ would be fixed, and the base value of $\boldsymbol{r}_{xy}$ was given. The coefficient of determination was varied by choosing values of *k* that yield R^2 equal to .2, .5, or .8. In essence, the value of R^2 was manipulated by inserting the quantity *k* in the equation $R^2 = \sqrt{k} * \boldsymbol{r}'_{xy} * \boldsymbol{R}^{-1}_{xx'} * \boldsymbol{r}_{xy} * \sqrt{k}$. In this equation, $\boldsymbol{r}_{xy}$ contains the initial values. The regression weights were estimated using $\sqrt{k}\boldsymbol{r}_{xy}$ as the correlations between the predictors and the criterion variable with *k* being chosen to yield the coefficient of determination .2, .5, or .8 depending on the experimental condition. The set of values, k_1, k_2, and k_3 were used to generate the EWEI model and the POD model.

For the EWEI model, the new values of $\boldsymbol{r}_{xy}$ for each group were the same because the regression coefficients are not different. For each value of the coefficient of determination, the values of $\boldsymbol{r}_{xy}$ were the same for the reference group and the focal group. The values of *k* were the same for the reference group and the focal

group. When the coefficient of determination was 0.2 for each group, the value of $k_1 = 1$ was applied. When the coefficient of determination was 0.5 for each group, the value of $k_2 = 1.58$ was applied. When the coefficient of determination was 0.8 for each group, the value of $k_3 = 2.01$ was applied.

For the portion of the study in which the generating model was the EWEI model, there were three different EWEI models, one in which $R^2 = .2$ for both the reference and focal group, one in which $R^2 = .5$ for both the reference and focal group, and one in which $R^2 = .8$ for both the reference and focal group. Under each of these three conditions, the β weights were : (0.0486, 0.0486, 0.2514,0.2514), (0.077, 0.077, 0.397, 0.397) and (0.097, 0.097, 0.5028, 0.5028) corresponding to the values of k_1=1, $k_2 = 1.58$, and $k_3 = 2.01$ respectively.

POD Generating Model. For the POD model, the new values of $\boldsymbol{r}_{xy}$ were different for the reference and focal groups because the regression coefficients were proportional to each other. The value of *k* for the reference group was the same as the one computed by using the corresponding value of the coefficient of determination. The value of *k* for the focal group was chosen from two other conditions.

If the coefficient of determination was 0.2 to construct the new values of $\boldsymbol{r}_{xy}$, the value of $k_1 = 1$ was used for the reference group. A value different from k_1 was used for the focal group. An important feature of the POD model is this: if the predictor inter-correlation matrix $\boldsymbol{R}_{xx}$ is held constant, then a reference and focal group with unequal $\boldsymbol{r}_{xy}$ matrices will have unequal coefficients of determination.

Thus, in the study with the POD model as the generating model, the reference and focal groups have unequal coefficients of determination (except when $k_g = -1 \; or \; 1$).

In the sub-study with the POD as the generating model, there were three POD models: 1) one with R^2 = .2 for the reference group and R^2 = .5 for the focal group, 2) one with R^2 = .5 for the reference group and R^2 = .8 for the focal group, and 3) one with R^2 = .8 for the reference group and R^2 = .2 for the focal group. Within each reference or focal group, the value of R^2 was set by manipulating the value of k in the equation $R^2 = \sqrt{k} * \boldsymbol{r}'_{xy} * \boldsymbol{R}^{-1}_{xx'} * \boldsymbol{r}_{xy} * \sqrt{k}$. For the first condition, the β weights were (0.049, 0.049, 0.251, 0.251) for the reference group, and (0.077, 0.077, 0.397,0 .397) for the focal group. For the second condition, the β weights were (0.077, 0.077, 0.397, 0.397) for the reference group, and (0.097, 0.097, 0.503, 0.503) for the focal group. For the third condition, the β weights were (0.097, 0.097, 0.503, 0.503) for the reference group, and (0.049, 0.049, 0.251, 0.251) for the focal group. For each of these conditions, the β weights for the focal and reference groups were related by a multiplicative constant as posited by the POD model. These three conditions correspond to the three possible combinations of the three coefficients of determination: (.2, .5), (.5, .8), and (.8, .2).

MMM Generating Model

When constructing the variance covariance matrices for each group for the MMM, I assumed that the values of the correlation between predictors and the response

variable in the two groups differ by an additive constant *k*. As with the POD model, if the predictor inter-correlation $\boldsymbol{R}_{\mathbf{xx}}$ is held constant, the reference and focal group have unequal coefficients of determination unless $k = 0$.

Because the regression coefficients differ by an additive constant between groups and the predictor covariance matrix is the same for both groups (a compound symmetric matrix), the criterion correlation matrix can be written as:

$$\boldsymbol{r}_{xy}^{*} = \boldsymbol{r}_{xy} + k * \boldsymbol{1}$$

Here, the column vector $\mathbf{r}_{\mathbf{xy}}$ represents the initial values described above, which were the same for each group. The values of $\boldsymbol{r}_{xy}^{*}$ differs by an additive constant between the two groups.

The regression coefficients can be expressed as: $\boldsymbol{\beta} = \boldsymbol{R}_{xx'}^{-1} * \boldsymbol{r}_{xy}$. If we insert the new values of $\boldsymbol{r}_{xy}^{*}$, which is defined above, the new regression coefficients can be written as the following:

$$\boldsymbol{\beta}^{*} = \boldsymbol{R}_{xx'}^{-1} * \boldsymbol{r}_{xy}^{*} = \boldsymbol{R}_{xx'}^{-1}(\boldsymbol{r}_{xy} + k * \boldsymbol{1}) = \boldsymbol{\beta} + \boldsymbol{R}_{xx'}^{-1} * k * \boldsymbol{1} = \boldsymbol{\beta} + k' * \boldsymbol{1}$$

To solve the values of *k*, I inserted $\boldsymbol{r}_{xy}^{*} = \boldsymbol{r}_{xy} + k * \boldsymbol{1}$ into the equation $\boldsymbol{\beta} = \boldsymbol{R}_{xx'}^{-1} * \boldsymbol{r}_{xy}$.

Thus, I got the following results:

$$R^2 = \boldsymbol{r}_{xy}' \boldsymbol{R}_{xx'}^{-1} \boldsymbol{r}_{xy}$$

$$= (\boldsymbol{r}_{xy}' + k * \boldsymbol{1}') \boldsymbol{R}_{xx'}^{-1} (\boldsymbol{r}_{xy} + \boldsymbol{1} * k)$$

$$= \boldsymbol{r}_{xy}' * \boldsymbol{R}_{xx}^{-1} * \boldsymbol{r}_{xy} + k * \boldsymbol{1}' * \boldsymbol{R}_{xx'}^{-1} \boldsymbol{r}_{xy} + \boldsymbol{r}_{xy}' * \boldsymbol{R}_{xx'}^{-1} * \boldsymbol{1} * k + k * \boldsymbol{1}' * \boldsymbol{R}_{xx'}^{-1} * \boldsymbol{1} * k$$

$$= \boldsymbol{R}^2 + k * \boldsymbol{1}' * \boldsymbol{\beta} + \boldsymbol{\beta}' * \boldsymbol{1} * k + k^2 * \boldsymbol{1}' * \boldsymbol{R}_{xx'}^{-1} * \boldsymbol{1}$$

The constants of the second term and third term, $\boldsymbol{1}' * \boldsymbol{\beta}$ and $\boldsymbol{\beta}' * \boldsymbol{1}$, can be simplified as the sum of the elements of the matrix $\boldsymbol{\beta} = \boldsymbol{R}_{xx'}^{-1} * \boldsymbol{r}_{xy}$. The constant of the fourth term, $\boldsymbol{1}' * \boldsymbol{R}_{xx'}^{-1} * \boldsymbol{1}$, can be simplified as the sum of the elements of the inverse of the $\boldsymbol{R}_{xx}$ matrix.

To solve the equation with respect to *k*, I inserted different values of *R* squared for each condition, that resulted in different *k* values for different groups. The values were notated as: $k_{m1} = 0$, $k_{m2} = 0.188$, and $k_{m3} = 0.320$.

In the sub-study with the MMM as the generating model, there were three conditions: one with $R^2 = .2$ for the reference group and $R^2 = .5$ for the focal group, one with $R^2 = .5$ for the reference group and $R^2 = .8$ for the focal group, and one with $R^2 = .8$ for the reference group and $R^2 = .2$ for the focal group. For the first condition, the β weights were (0.049, 0.049, 0.251, 0.251) for the reference group, and (0.148, 0.148, 0.350, 0.350) for the focal group. For the second condition, the β weights were (0.148, 0.148, 0.350, 0.350) for the reference group, and (0.217, 0.217, 0.420, 0.420) for the focal group. For the third condition, the β weights were (0.217, 0.217, 0.420, 0.420) for the reference group, and (0.049, 0.049, 0.251, 0.251) for the focal group. For each of these conditions, the β weights for the focal and reference groups were related by an additive constant as posited by the MMM. Again, these three conditions correspond to the three combinations of the coefficients of determination: (.2, .5), (.5, .8), and (.8, .2).

FMM Generating Model

For the FMMM, I used a similar principle as the EWEI and POD models. As stated above, by using the equation $R^2 = \boldsymbol{r}_{xy}^{'} * \boldsymbol{R}_{xx}^{-1} * \boldsymbol{r}_{xy}$, the different values of *R* squared (0.2, 0.5 or 0.8) yielded different values of $\boldsymbol{r}_{xy}$. The regression weights were computed by using $\boldsymbol{\beta} = \boldsymbol{R}_{xx'}^{-1} * \boldsymbol{r}_{xy}$.

For the reference group, the initial value of $\boldsymbol{r}_{xy}$ was set as (0.3, 0.3, 0.5, 0.5). For the focal group, there were two conditions.

Under the first condition, for the focal group, the value of $\boldsymbol{r}_{xy}$ was permuted as (0.3, 0.5, 0.5, 0.3). Under the second condition, for the focal group, the value of $\boldsymbol{r}_{xy}$ was increasing and set as (0.3, 0.4, 0.7, 0.75). The first condition simulated a condition in which the rank order of weights differs substantially from the reference to the focal group. The Spearman correlation for the rank of the weights was zero. The second condition simulated a condition in which the FMM model holds, but the rank order of weights was the same in both the reference and focal groups.

In the sub-study with the FMM as the generating model, there were three FMMs for each condition of $\boldsymbol{r}_{xy}$ in the focal group: one with R^2 = .2 for the reference group and R^2 = .5 for the focal group, one with R^2 = .8 for the reference group and R^2 = .2 for the focal group, and one with R^2 = .5 for the reference group and R^2 = .8 for the focal group.

Under the condition that the $\boldsymbol{r}_{xy}$ in the focal group was set as (0.3, 0.3, 0.5, 0.5), the regression weights for the three FMMs were computed. With R^2 = .2 for the

reference group and R^2 = .5 for the focal group, the β weights were (0.049, 0.049, 0.251, 0.251) for the reference group, and (0.077, 0.397, 0.077, 0.397) for the focal group. With R^2 = .5 for the reference group and R^2 = .8 for the focal group, the β weights were (0.077, 0.077, 0.397, 0.397) for the reference group, and (0.097, 0.503, 0.503, 0.097) for the focal group. With R^2 = .8 for the reference group and R^2 = .2 for the focal group, the β weights were (0.097, 0.097, 0.053, 0.053) for the reference group, and (0.049, 0.251, 0.251, 0.049) for the focal group.

Under the condition that the $\boldsymbol{r}_{xy}$ in the focal group was set as (0.3, 0.4, 0.7, 0.75), the regression weights for the three FMMs were computed. With R^2 = .2 for the reference group and R^2 = .5 for the focal group, the β weights were (0.049, 0.049, 0251, 0.251) for the reference group, and (-0.043, 0.068, 0.402, 0.458) for the focal group. With R^2 = .5 for the reference group and R^2 = .8 for the focal group, the β weights were (0.077, 0.077, 0.397, 0.397) for the reference group, and (-0.055, 0.086, 0.509, 0.580) for the focal group. With R^2 = .8 for the reference group and R^2 = .2 for the focal group, the β weights were (0.097, 0.097, 0.503, 0.503) for the reference group, and (-0.028, 0.043, 0.255, 0.290) for the focal group.

Table 6 summarizes the regression weights for each generating model stated previously.

Table 6

The R Squared Values and the Regression Coefficients for Each Model

Generating Model	*R* Squared for the Reference Group	*R* Squared for the Focal Group	Regression Coefficients for the Reference Group	Regression Coefficients for the Focal Group
EWEI	0.2	0.2	(0.0486, 0.0486, 0.2514,0.2514)	(0.0486, 0.0486, 0.2514,0.2514)
	0.5	0.5	(0.077, 0.077, 0.397, 0.397)	(0.077, 0.077, 0.397, 0.397)
	0.8	0.8	(0.097, 0.097, 0.5028, 0.5028)	(0.097, 0.097, 0.5028, 0.5028)
EWUI	0.2	0.2	(0.0486, 0.0486, 0.2514,0.2514)	(0.0486, 0.0486, 0.2514,0.2514)
	0.5	0.5	(0.077, 0.077, 0.397, 0.397)	(0.077, 0.077, 0.397, 0.397)
	0.8	0.8	(0.097, 0.097, 0.5028, 0.5028)	(0.097, 0.097, 0.5028, 0.5028)
PODUI	0.2	0.5	(0.049, 0.049, 0.251, 0.251)	(0.077, 0.077, 0.397,0 .397)
	0.5	0.8	(0.077, 0.077, 0.397,0 .397)	(0.097, 0.097, 0.503, 0.503)
	0.8	0.2	(0.097, 0.097, 0.503, 0.503)	(0.049, 0.049, 0.251, 0.251)
MMMUI	0.2	0.5	(0.049, 0.049, 0.251, 0.251)	(0.148, 0.148, 0.350, 0.350)
	0.5	0.8	(0.148, 0.148, 0.350, 0.350)	(0.217, 0.217, 0.420, 0.420)
	0.8	0.2	(0.217, 0.217, 0.420, 0.420)	(0.049, 0.049, 0.251, 0.251)
FMMUI with Different Rank Orders	0.2	0.5	(0.049, 0.049, 0.251, 0.251)	(0.077, 0.397, 0.077, 0.397)
	0.5	0.8	(0.077, 0.077, 0.397, 0.397)	(0.097, 0.503, 0.503, 0.097)
	0.8	0.2	(0.097, 0.097, 0.053, 0.053)	(0.049, 0.251, 0.251, 0.049)

Generating Model	*R* Squared for the Reference Group	*R* Squared for the Focal Group	Regression Coefficients for the Reference Group	Regression Coefficients for the Focal Group
FMMUI with the Same Rank Order	0.2	0.5	(0.049, 0.049, 0251, 0.251)	(-0.043, 0.068, 0.402, 0.458)
	0.5	0.8	(0.077, 0.077, 0.397, 0.397)	(-0.055, 0.086, 0.509, 0.580)
	0.8	0.2	(0.097, 0.097, 0.503, 0.503)	(-0.028, 0.043, 0.255, 0.290)

Construction of different intercepts for different groups

In this study, the intercepts for the reference group and the focal group were different in the unequal intercept conditions. They differed by 0.2 standard deviation of *Y*. The intercept was 0 in the reference group. For all six models, when the data was generated, I used the same matrices as mentioned above for models with equal intercepts. Then, the values of *Y* were computed separately for each group.

In general, I computed the regression weights in an indirect way. The correlations between *X* and *Y* were the same for the EWEI and EWUI model for the reference group and the focal group. The correlations between *X* and *Y* were different for the PODUI, MMMUI and FMMUI for the reference group and the focal group. For each model, if the correlation between *X* and *Y* was computed, the variance covariance matrix $\boldsymbol{R}_{xy}$ was computed for the reference group and the focal group. Here $\boldsymbol{R}_{xy}$ is a 5x5 matrix containing correlations of predictors with the criterion as well as the intercorrelations of the predictors. The MVNORM function in R was used to generate the data. The function MVNORM generates samples with sample size *n*, a given

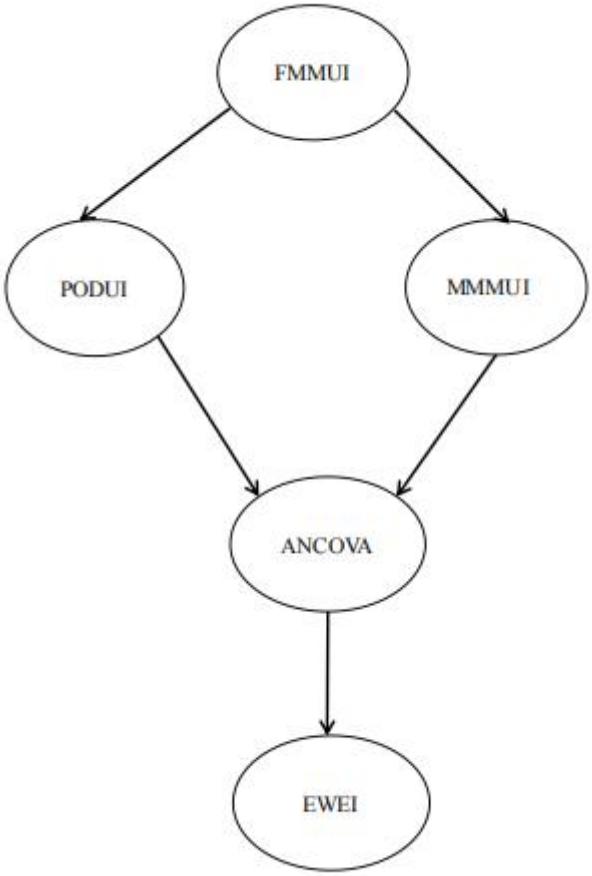
mean value and a given covariance matrix. For each replication, MVNORM generates a vector containing a criterion score Y and four predictor values X_1 - X_4 that are drawn from a multivariate normal distribution with population covariance matrix $\boldsymbol{R}_{xy}$. It uses an eigen decomposition to generate the random numbers.

Hypothesis Tests

As described in the literature review, the ten models have a hierarchical structure. Figure 3 shows the nested structure in the analysis.

Figure 3

The Nested Structure of Models Used for the Analysis



For the six models used in this study, the EWEI model is nested in all of the other models. The EWUI model is nested in the PODUI, the MMMUI and the FMMUI. The PODUI model is nested in the FMMUI. The MMMUI is nested in the FMMUI. For all models, the null hypothesis can be written as: $R_G^2 = R_F^2$, where G represents the generating model and F represents the fitted model. When the generating model is nested in the fitted model, the Type I error rate is reported. When the fitted model is nested in the generating model, the power is reported. If the models

being compared had the same number of parameters, the *F* statistic could not be computed, so power and Type I errors were not computed (e.g. POD and MMM).

Measures

In this section, the independent variables and dependent variables in the experiment are described.

Independent Variables

The study was composed of six sub-studies. Each sub-study had a different generating model: EWEI, EWUI (ANCOVA), PODUI, MMMUI, or FMMUI with two variations of the FMMUI model. Within each sub-study, the independent variables were the five fitted models defined previously, total sample size and different values of *R* squared. For the total sample size of each group, I used the values 100, 500 and 1000.

Dependent Variables

The Type I error and power were reported for different fitted models within generating models as described in the section of "Hypothesis Tests."

Practical issue

Some practical issues are related to the selection of models. If we have a specific research question and know which model to use, two models will be compared. The supposed model and the model with fewer predictors which is close to the supposed model will be used. The methods of model selection mentioned above will provide evidence for the model selection.

If one of the models is the hypothesis, then all five models should be fitted. The model as the hypothesis should be compared with all the other models simultaneously. Then, the simplest model that is better than everything below it will be picked. In what follows, the generating model as the hypothesis is said to be "identified" when the following two things happen: First, all models that are special cases of the generating model were rejected at the 0.05 level. Second, the generating model hypothesis was retained at 0.05 level when compared to all models that include it (the generating model) as a special case. At the end of the results below, I will report data on the number of times the generating model can be identified when compared to all of the other models (Table 22). First, however, I will report the Type I error rate and power results when the generating model is compared to one of the other models using the *F* statistic.

In Table 22, under each condition of different sample sizes, I computed the total number of rejections. If the generating model was the EWEI model, I reported the total number of replications in which the null hypothesis was retained for all the other models as the alternative. If the generating model was the EWUI model, I reported the total number of replications in which the EWEI was rejected with respect to at least one other model. If the generating model was the PODUI model, I reported the total number of replications in which the EWEI and the EWUI null hypotheses were rejected when compared to the PODUI model, and the PODUI model null hypothesis was retained when compared to the FMMUI. If the generating model was the

MMMUI, I reported the total number of replications in which the EWEI and EWUI null hypotheses were retained when compared to the MMMUI, and the MMMUI null hypothesis was retained when compared to the FMMUI. If the generating model was the FMMUI, I reported the total number of replications in which the FMMUI fit significantly better than all the other models.

The method for picking the "optimal" model, using the principles mentioned in this section, was applied to Park's (2018) study. In the literature review, I summarized two families of analyses. The first set of Park's analyses included 13 coursework covariates and a categorical variable "ethnicity." The second set of analyses included the 13 coursework covariates and an additional covariate "prior achievement," and a categorical variable "ethnicity." Park fitted all ten models and provided the results of model comparisons. I apply the principles in this section of model identification to the result of Park's study.

Results

In this section, the EWEI model, the ANCOVA (EWUI) model, the PODUI model, the MMMUI, and the FMMUI (under two conditions) are set as the generating and fitted models. From Table 7 to Table 21, the tables present the Type I error rate, the power, and the relevant statistics such as the values of *R* squared, adjusted *R* squared and other statistics. In this section, for the Type I error rate, the significance level is 5%. In educational studies, when 5% is set as the significance level, the Type I error rate is considered as acceptable if the actual rate is less than or equal to 7.5% (Blanca

et al., 2018). The authors also provided the lower bound. The value is 2.5%. In this study, I focus on the upper bound. That is: the Type I error rate should not be greater than 5%. The value greater than 0.8 is set as the standard for power.

Throughout Table 7 to Table 21, tables with *R* squared and adjusted *R* squared values, including Table 8, Table 11, Table 14, Table 17, Table 19, and Table 21, are used to verify the correctness of the simulation. They are used to verify that the *R* squared values of the fitted models are reasonable. For each generating model, the *R* squared values for the fitted models should be close to and a little higher than the average value of the generating *R* squared values mentioned in Table 5 in the method part. Table 7, Table 10, Table 13, and Table 16 are used to answer the first research question. Table 9, Table 12, Table 15, Table 18, and Table 20 are used to answer the second research question.

Table 22 is used to answer the third research question: is the generating model likely to be identified when the generating model is compared with all of the other models to which it is hierarchically related?

Table 7

The Type I Error Rate When the Generating Model is the EWEI Model

Sample Size	Comparison Model	Generating *R* squared		
		Reference Group 0.2 Focal Group 0.2	Reference Group 0.5 Focal Group 0.5	Reference Group 0.8 Focal Group 0.8
100	EWUI	7.2%	9.3%	8.8%
	PODUI	3.5%	4.3%	4.7%
	MMMUI	3.5%	4.5%	4.2%
	FMMUI	3.5%	3.0%	3.4%
500	EWUI	6.4%	5.7%	5.9%
	PODUI	3.8%	2.6%	3.3%
	MMMUI	3.6%	3.7%	3.5%
	FMMUI	3.7%	2.6%	2.6%
1000	EWUI	5.9%	6.3%	5.9%
	PODUI	3.3%	3.6%	3.4%
	MMMUI	3.1%	2.8%	3.1%
	FMMUI	2.9%	1.9%	2.4%

Table 8

Values of R Squared, Adjusted R Squared, and the F Statistic When the Generating Model is the EWEI Model

Sample Size	Generating *R* Squared Reference	Generating *R* Squared Focal	Fitted Model	Fitted *R* Squared	Fitted Adjusted *R* Squared	*F* Statistic
100	0.2	0.2	EWEI	0.212	0.197	13.43
			EWUI	0.217	0.197	10.96
			PODUI	0.221	0.197	9.32
			MMMUI	0.221	0.197	9.31
			FMMUI	0.233	0.197	9.54
	0.5	0.5	EWEI	0.592	0.584	72.28
			EWUI	0.595	0.584	58.09
			PODUI	0.597	0.584	48.58
			MMMUI	0.597	0.584	48.61
			FMMUI	0.603	0.584	32.72
	0.8	0.8	EWEI	0.802	0.798	201.21
			EWUI	0.803	0.798	161.32
			PODUI	0.804	0.798	134.63
			MMMUI	0.804	0.798	134.65
			FMMUI	0.806	0.798	90.08
500	0.2	0.2	EWEI	0.203	0.200	63.72
			EWUI	0.204	0.200	51.19
			PODUI	0.205	0.200	42.78
			MMMUI	0.205	0.200	42.83
			FMMUI	0.207	0.200	28.89
	0.5	0.5	EWEI	0.587	0.584	354.24
			EWUI	0.587	0.585	283.57
			PODUI	0.587	0.585	236.71
			MMMUI	0.587	0.585	236.49
			FMMUI	0.589	0.585	157.99
	0.8	0.8	EWEI	0.799	0.799	977.03
			EWUI	0.800	0.799	797.80
			PODUI	0.800	0.799	665.00
			MMMUI	0.800	0.799	665.04
			FMMUI	0.800	0.799	443.68

Sample Size	Generating *R* Squared Reference	Generating *R* Squared Focal	Fitted Model	Fitted *R* Squared	Fitted Adjusted *R* Squared	*F* Statistic
1000	0.2	0.2	EWEI	0.201	0.200	125.97
			EWUI	0.202	0.200	100.97
			PODUI	0.202	0.200	84.31
			MMMUI	0.202	0.200	84.31
			FMMUI	0.203	0.200	56.53
	0.5	0.5	EWEI	0.587	0.586	709.35
			EWUI	0.587	0.586	567.67
			PODUI	0.587	0.586	473.21
			MMMUI	0.587	0.586	473.23
			FMMUI	0.588	0.586	315.81
	0.8	0.8	EWEI	0.800	0.799	1996.31
			EWUI	0.800	0.799	1997.23
			PODUI	0.800	0.799	1331.16
			MMMUI	0.800	0.799	1331.20
			FMMUI	0.800	0.799	887.73

Table 7 presents the Type I error rate when the generating model is the EWEI model. When the comparison model is the EWUI model and the sample size is 100, the Type I error rate inflates to a value between 8% and 9%. When the comparison model is the EWUI model and the sample size is 500 or 1000, the Type I error rate is around 6%. When the comparison models are other models, the Type error rates are lower than 5%. Since the Type I error rate is based on the *F* statistic, the degrees of freedom impact the Type I error rate. The difference of degrees of freedom of residuals between the EWEI model and the EWUI model is 1, smaller than that value between the EWEI model and other models. Thus, if the generating model is the EWEI model and the fitted model is the EWUI model, the Type I error rate inflates.

The sample size is another factor impacting the Type I error rate. Under each condition, when the fitted model is the EWUI model, the smaller sample size leads to a larger Type I error rate.

Table 8 presents the values of *R* squared, adjusted *R* squared, and other statistics for each fitted model. At each sample size, when the *R* squared value is 0.2 for the reference group for the generating model, the *R* squared values for all the fitted models are around 0.2. When the *R* squared value is 0.8 for the generating model, the *R* squared values for all the fitted models are around 0.8.

In Table 8, when the *R* squared value is 0.5 for the generating model, the *R* squared values for all the fitted models are around 0.6, higher than the expected value 0.5. Since this value is higher for all of the models, the result of the Type I error rate will not be influenced.

Table 9

The Power When the Generating Model is the EWUI Model

Sample Size	Comparison Model	Generating *R* squared		
		Reference Group 0.2 Focal Group 0.2	Reference Group 0.5 Focal Group 0.5	Reference Group 0.8 Focal Group 0.8
100	EWEI	39.4%	59.9%	88.2%
500	EWEI	94.6%	99.8%	100%
1000	EWEI	99.9%	100%	100%

Table 10

The Type I error rate When the Generating Model is the EWUI Model

Sample Size	Comparison Model	Generating *R* squared		
		Reference Group 0.2 Focal Group 0.2	Reference Group 0.5 Focal Group 0.5	Reference Group 0.8 Focal Group 0.8
100	PODUI	7.7%	6.5%	6.2%
	MMMUI	7.1%	6.9%	5.8%
	FMMUI	4.2%	4.0%	3.8%
500	PODUI	5.1%	6.0%	6.8%
	MMMUI	5.9%	7.1%	7.3%
	FMMUI	4.3%	4.1%	4.5%
1000	PODUI	5.9%	6.8%	5.9%
	MMMUI	7.1%	5.8%	5.8%
	FMMUI	3.4%	4.2%	3.2%

Table 11

Values of R Squared, Adjusted R Squared, and the F Statistic When the Generating Model is the EWUI Model

Sample Size	Generating *R* Squared Reference	Generating *R* Squared Focal	Fitted Model	Fitted *R* Squared	Fitted Adjusted *R* Squared	*F* Statistic
100	0.2	0.2	EWEI	0.211	0.195	13.28
			EWUI	0.225	0.205	11.46
			PODUI	0.229	0.205	9.72
			MMMUI	0.229	0.205	9.72
			FMMUI	0.241	0.205	6.81
	0.5	0.5	EWEI	0.587	0.578	70.63
			EWUI	0.599	0.588	59.08
			PODUI	0.601	0.588	49.40
			MMMUI	0.601	0.588	49.43
			FMMUI	0.607	0.588	34.27
	0.8	0.8	EWEI	0.794	0.790	191.77
			EWUI	0.805	0.800	163.36
			PODUI	0.805	0.800	136.30
			MMMUI	0.805	0.800	136.35
			FMMUI	0.809	0.800	91.22
500	0.2	0.2	EWEI	0.201	0.199	62.96
			EWUI	0.211	0.208	53.61
			PODUI	0.213	0.208	44.83
			MMMUI	0.213	0.208	44.85
			FMMUI	0.215	0.208	30.24
	0.5	0.5	EWEI	0.581	0.579	345.85
			EWUI	0.591	0.589	288.35
			PODUI	0.591	0.589	240.46
			MMMUI	0.591	0.589	240.47
			FMMUI	0.593	0.589	160.64
	0.8	0.8	EWEI	0.792	0.791	549.59
			EWUI	0.802	0.801	807.82
			PODUI	0.802	0.801	673.31
			MMMUI	0.802	0.801	673.39
			FMMUI	0.803	0.801	499.25

Models of Moderation in Regression

Sample Size	Generating *R* Squared Reference	Generating *R* Squared Focal	Fitted Model	Fitted *R* Squared	Fitted Adjusted *R* Squared	*F* Statistic
1000	0.2	0.2	EWEI	0.199	0.198	124.45
			EWUI	0.210	0.208	106.11
			PODUI	0.210	0.208	88.59
			MMMUI	0.210	0.208	88.60
			FMMUI	0.211	0.208	59.39
	0.5	0.5	EWEI	0.581	0.580	692.58
			EWUI	0.591	0.590	577.29
			PODUI	0.591	0.590	481.22
			MMMUI	0.591	0.590	481.25
			FMMUI	0.592	0.590	321.15
	0.8	0.8	EWEI	0.792	0.791	1901.20
			EWUI	0.802	0.801	1616.85
			PODUI	0.802	0.801	1347.48
			MMMUI	0.802	0.801	1347.55
			FMMUI	0.802	0.801	898.64

Table 9 presents the power when the generating model is the EWUI model and the comparison model is the EWEI model. When the sample size is 100 and the *R* squared value is 0.2, the power is 39.4%, the lowest one. Under the condition that the sample size is 100, the power increases when the generating *R* squared value is larger. Under the conditions that the sample size is 500 or 1000, all of the values of the power are larger than 90%. Higher power is related to a larger sample size (500) or a larger *R* squared value for the generating model (0.8).

Table 10 presents the Type I error rate when the generating model is the EWUI model and the comparison models are the PODUI model, the MMMUI, and the FMMUI. When the comparison model is the PODUI model, under each condition, the Type I error rate is around 6%. When the comparison model is the MMMUI, under

each condition, the Type I error rate is around 6% to 7%. When the comparison model is the FMMUI, under each condition, the Type I error rate is around 4%. As stated at the beginning of this section, 7.5% is an acceptable upper bound in the educational studies for the Type I error rate. Thus, under most conditions, the Type I error rates are reasonable. The smaller sample size and the smaller difference of the degrees of freedom of the residuals between the generating model and comparison model are the factors related to the larger Type I error rate. For example, when the sample size is 100, and the comparison model is the PODUI model, the Type I error rate inflates to 7.7%. For the FMMUI as the comparison model, since the difference of the degrees of freedom of the residuals is lager, under each condition, the Type I error rate decreases.

Table 11 presents the values of *R* squared, adjusted *R* squared and other statistics for all the fitted models when the generating model is the EWUI model. Similar to the result in Table 8, when the *R* squared value is 0.5 for the generating model, the *R* squared values for all the fitted models are around 0.6. That will not impact the results in Table 9 and Table 10, since all of the fitted models have the higher *R* squared values.

Table 12

The Power When the Generating Model is the PODUI Model

Sample Size	Comparison Model	Generating *R* squared		
		Reference Group 0.2 Focal Group 0.5	Reference Group 0.5 Focal Group 0.8	Reference Group 0.8 Focal Group 0.2
100	EWEI	62.7%	75.3%	98.0%
	EWUI	60.6%	59.3%	98.9%
500	EWEI	100%	100%	100%
	EWUI	100%	100%	100%
1000	EWEI	100%	100%	100%
	EWUI	100%	100%	100%

Table 13

The Type I error rate When the Generating Model is the PODUI Model

Sample Size	Comparison Model	Generating *R* squared		
		Reference Group 0.2 Focal Group 0.5	Reference Group 0.5 Focal Group 0.8	Reference Group 0.8 Focal Group 0.2
100	FMMUI	8.9%	8.1%	13.5 %
500	FMMUI	7.3%	8.4%	13.7 %
1000	FMMUI	8.2%	8.2%	13.6%

Table 14

Values of R Squared, Adjusted R Squared, and the F Statistic When the Generating Model is the PODUI Model

Sample Size	Generating *R* Squared Reference	Generating *R* Squared Focal	Fitted Model	Fitted *R* Squared	Fitted Adjusted *R* Squared	*F* Statistic
100	0.2	0.5	EWEI	0.339	0.325	25.50
			EWUI	0.351	0.335	21.50
			PODUI	0.370	0.350	19.27
			MMMUI	0.369	0.349	19.18
			FMMUI	0.381	0.351	13.25
	0.5	0.8	EWEI	0.638	0.630	87.70
			EWUI	0.649	0.640	73.43
			PODUI	0.659	0.648	63.96
			MMMUI	0.658	0.648	63.36
			FMMUI	0.665	0.649	42.79
	0.8	0.2	EWEI	0.455	0.444	41.81
			EWUI	0.468	0.454	34.96
			PODUI	0.516	0.501	35.08
			MMMUI	0.512	0.497	34.48
			FMMUI	0.527	0.504	23.96

Models of Moderation in Regression

Sample Size	Generating *R* Squared Reference	Generating *R* Squared Focal	Fitted Model	Fitted *R* Squared	Fitted Adjusted *R* Squared	*F* Statistic
500	0.2	0.5	EWEI	0.330	0.327	122.95
			EWUI	0.340	0.337	103.08
			PODUI	0.357	0.354	92.43
			MMMUI	0.355	0.351	91.48
			FMMUI	0.360	0.354	62.01
	0.5	0.8	EWEI	0.633	0.632	431.69
			EWUI	0.644	0.642	360.77
			PODUI	0.653	0.651	312.25
			MMMUI	0.651	0.650	310.62
			FMMUI	0.654	0.651	208.64
	0.8	0.2	EWEI	0.446	0.443	200.83
			EWUI	0.456	0.453	167.41
			PODUI	0.506	0.503	169.89
			MMMUI	0.499	0.496	165.62
			FMMUI	0.508	0.503	113.85
1000	0.2	0.5	EWEI	0.330	0.328	245.94
			EWUI	0.340	0.338	205.86
			PODUI	0.357	0.355	184.61
			MMMUI	0.355	0.353	182.81
			FMMUI	0.358	0.355	123.47
	0.5	0.8	EWEI	0.634	0.633	865.49
			EWUI	0.644	0.643	722.92
			PODUI	0.653	0.652	625.92
			MMMUI	0.652	0.651	625.66
			FMMUI	0.654	0.652	417.14
	0.8	0.2	EWEI	0.446	0.445	402.74
			EWUI	0.456	0.455	335.46
			PODUI	0.506	0.504	340.31
			MMMUI	0.499	0.497	331.16
			FMMUI	0.507	0.504	227.45

Table 12 and Table 13 present the power and Type I error rate when the generating model is the PODUI model. Table 14 presents the results of the values of the *R* squared, adjusted *R* squared and other statistics.

Table 14 is used to verify if the *R* squared values are reasonable for the fitted models. When the *R* squared values are 0.2 for the reference group and 0.5 for the focal group, the *R* squared values for all the fitted models are around 0.35, the average value of 0.2 and 0.5. When the *R* squared values are 0.5 for the reference group and 0.8 for the reference group, the *R* squared values for all the fitted models are around 0.65, the average value of 0.5 and 0.8. When the *R* squared values are 0.8 for the reference group and 0.2 for the focal group, the *R* squared values for all the fitted models are around 0.5, the average value of 0.8 and 0.2.

Table 12 presents the power when the generating model is the PODUI model, and the comparison models are the EWEI model and the EWUI model. Generally, when the sample size is 100 and the *R* squared value for the reference group is 0.2 or 0.5, the power is lower than 80%.

When the comparison model is the EWEI model, under most conditions, the power is larger than 90%. When the generating *R* squared value for the reference group is 0.2 or 0.5, and the sample size is 100, the power decreases. When the *R* squared value of the reference group is 0.8 and the difference of the *R* squared values between the reference group and the focal group is 0.6, the power is around 99%. When the generating *R* squared values is 0.2 for the reference group, the fitted *R* squared values are 0.35 for the generating model and the comparison model, the smaller sample size (100) and the lower value of the fitted *R* squared values lead to the lower power. For other conditions when sample size is 100, the larger fitted R

square values are not related to the higher power. Table 14 shows that under the condition that the *R* squared values of the fitted models are 0.65 and the sample size is 100, the difference of the fitted *R* squared values between the EWEI model and the PODUI model is 0.021. Table 14 shows that under the condition that the R squared values of the fitted models are about 0.5 and the sample size is 100, the difference of the fitted *R* squared values between the EWEI model and the PODUI model is 0.057. That explains the reason that the power increases when the *R* squared values of the fitted models decrease under the condition that the sample size is 100. Under the condition that the sample size is 100, when the difference of the *R* squared values between the reference group and the focal group is larger for the generating model, the power increases.

When the comparison model is the EWUI model, under most conditions, the power is larger than 90%. Under the condition that the sample size is 100, and the *R* squared value for the reference group is 0.2 or 0.5, the power is around 60%. When the *R* squared value of the reference group is 0.8 and the difference of the *R* squared values between the reference group and the focal group is 0.6, the power is around 99%. When the generating *R* squared values is 0.2 for the reference group, the fitted *R* squared values are 0.35 for the generating model and the comparison model, the smaller sample size (100) and the lower value of the fitted *R* squared values lead to the lower power. For other conditions when the sample size is 100, the larger fitted *R* square values are not related to the higher power. Table 14 shows that when the

sample size is 100 and the EWUI model is the comparison model, under the condition that the *R* squared values of the fitted model are about 0.65, the power is 60%, smaller than the power 98%, under the condition that the *R* squared values of the fitted models are about 0.5. Table 14 shows that under the condition that the *R* squared values of the fitted models are about 0.65 and the sample size is 100, the difference of the fitted *R* squared values between the EWUI model and the PODUI model is 0.01. Table 14 shows that under the condition that the *R* squared values of the fitted models are 0.5 and the sample size is 100, the difference of the fitted *R* squared values between the EWUI model and the PODUI model is 0.048. That explains the reason that the power increases when the *R* squared value of the fitted model decreases under the condition that the sample size is 100. Under the condition that the sample size is 100, when the difference of the *R* squared values between the reference group and the focal group is larger for the generating model, the power increases.

Table 13 presents the Type I error rate when the generating model is the PODUI model and the comparison models is the FMMUI. Under each condition, the Type I error rate is greater than 0.05. When the generating model is the PODUI model and the comparison model is the FMMUI, the null hypothesis of the PODUI model as the generating model is not retained at 0.05 level. One of the possible reason is: in the process of fitting the PODUI model, the iterative method is used. That may lead to the biased estimation of the *R* squared value of the PODUI model.

Table 15

The Power When the Generating Model is the MMMUI

Sample Size	Comparison Model	Generating *R* squared		
		Reference Group 0.2 Focal Group 0.5	Reference Group 0.5 Focal Group 0.8	Reference Group 0.8 Focal Group 0.2
100	EWEI	67.8%	77.6%	99%
	EWUI	69.7%	63.5%	99.4%
500	EWEI	100%	100%	100%
	EWUI	100%	99.9%	100%
1000	EWEI	100%	100%	100%
	EWUI	100%	100%	100%

Table 16

The Type I error rate When the Generating Model is the MMMUI

Sample Size	Comparison Model	Generating *R* squared		
		Reference Group 0.2 Focal Group 0.5	Reference Group 0.5 Focal Group 0.8	Reference Group 0.8 Focal Group 0.2
100	FMMUI	6.4%	6.7%	6.3%
500	FMMUI	5.2%	5.8%	6.7%
1000	FMMUI	5.8%	5.9%	4.8%

Table 17

Values of R Squared, Adjusted R Squared, and the F Statistic When the Generating Model is the MMMUI

Sample Size	Generating *R* Squared Reference	Generating *R* Squared Focal	Fitted Model	Fitted *R* Squared	Fitted Adjusted *R* Squared	*F* Statistic
100	0.2	0.5	EWEI	0.337	0.324	25.35
			EWUI	0.350	0.334	21.38
			PODUI	0.369	0.349	19.21
			MMMUI	0.372	0.352	19.43
			FMMUI	0.381	0.352	13.29
	0.5	0.8	EWEI	0.638	0.631	87.88
			EWUI	0.650	0.641	73.61
			PODUI	0.660	0.649	63.67
			MMMUI	0.660	0.650	63.95
			FMMUI	0.666	0.650	42.98
	0.8	0.2	EWEI	0.452	0.440	41.17
			EWUI	0.464	0.450	34.43
			PODUI	0.513	0.450	34.76
			MMMUI	0.519	0.504	35.44
			FMMUI	0.527	0.504	23.97
500	0.2	0.5	EWEI	0.329	0.326	122.50
			EWUI	0.340	0.336	102.72
			PODUI	0.357	0.353	92.23
			MMMUI	0.359	0.355	92.86
			FMMUI	0.360	0.355	62.25
	0.5	0.8	EWEI	0.634	0.633	433.11
			EWUI	0.644	0.643	361.97
			PODUI	0.653	0.651	313.13
			MMMUI	0.654	0.652	313.81
			FMMUI	0.655	0.652	209.57
	0.8	0.2	EWEI	0.441	0.439	197.65
			EWUI	0.452	0.449	164.76
			PODUI	0.502	0.499	167.83
			MMMUI	0.506	0.503	170.38
			FMMUI	0.508	0.503	113.93

Sample Size	Generating *R* Squared Reference	Generating *R* Squared Focal	Fitted Model	Fitted *R* Squared	Fitted Adjusted *R* Squared	*F* Statistic
1000	0.2	0.5	EWEI	0.329	0.328	244.86
			EWUI	0.339	0.337	204.98
			PODUI	0.356	0.354	184.11
			MMMUI	0.358	0.356	185.44
			FMMUI	0.359	0.356	123.97
	0.5	0.8	EWEI	0.635	0.634	868.18
			EWUI	0.645	0.644	725.20
			PODUI	0.653	0.652	627.66
			MMMUI	0.654	0.653	628.98
			FMMUI	0.655	0.653	419.67
	0.8	0.2	EWEI	0.442	0.441	396.59
			EWUI	0.452	0.451	330.35
			PODUI	0.503	0.501	336.01
			MMMUI	0.506	0.504	340.89
			FMMUI	0.507	0.504	227.58

Table 15 and Table 16 present the power and Type I error rate when the generating model is the MMMUI. Generally, when the sample size is 100 and the *R* squared value is for the reference group is 0.2 or 0.5, the power is lower than 80%.

Table 17 is used to verify if the *R* squared values are reasonable for the fitted models. Table 17 presents the results of the values of the *R* squared, adjusted *R* squared and other statistics when the generating model is the MMMUI. When the *R* squared values are 0.2 for the reference group and 0.5 for the focal group, the *R* squared values for all the fitted models are around 0.35, the average value of 0.2 and 0.5. When the *R* squared values are 0.5 for the reference group and 0.8 for the reference group, the *R* squared values for all the fitted models are around 0.65, the average value of 0.5 and 0.8. When the *R* squared values are 0.8 for the reference

group and 0.2 for the focal group, the *R* squared values for all the fitted models are around 0.5, the average value of 0.8 and 0.2.

Table 15 presents the power when the generating model is the MMMUI and the comparison models are the EWEI model and the EWUI model.

When the comparison model is the EWEI model, under most conditions, the power is larger than 90%. When the generating *R* squared value for the reference group is 0.2 or 0.5, and the sample size is 100, the power decreases. When the *R* squared value of the reference group is 0.8 and the difference of the *R* squared values between the reference group and the focal group is 0.6, the power is around 99%. When the generating *R* squared values is 0.2 for the reference group, the fitted *R* squared values are 0.35 for the generating model and the comparison model, the smaller sample size (100) and the lower value of the fitted *R* squared values leads to the lower power. For other conditions when the sample size is 100, the larger fitted *R* square values are not related to the higher power. Table 17 shows that under the condition that the *R* squared values of the fitted models are 0.65, and the sample size is 100, the difference of the fitted *R* squared values between the EWEI model and the MMMUI is 0.022. Table 11 shows that under the condition that the *R* squared values of the fitted models are 0.5 and the sample size is 100, the difference of the fitted *R* squared values between the EWEI model and the MMMUI is 0.067. That explains the reason that the power increases when the *R* squared values of the fitted models decrease under the condition that the sample size is 100. Under the condition that the

sample size is 100, when the difference of the *R* squared values between the reference group and the focal group is larger for the generating model, the power increases.

When the fitted model is the EWUI model, under most conditions, the power is larger than 90%. When the generating *R* squared value for the reference group is 0.2 or 0.5, and the sample size is 100, the power decreases. When the *R* squared value of the reference group is 0.8 and the difference of the *R* squared values between the reference group and the focal group is 0.6, the power is around 99%. When the generating *R* squared values is 0.2 for the reference group, the fitted *R* squared values are 0.35 for the generating model and the comparison model, the smaller sample size (100) and the lower value of the fitted *R* squared values lead to the lower power. For other conditions when the sample size is 100, the larger fitted *R* square values are not related to the higher power. Table 17 shows that under the condition that the *R* squared values of the fitted models are 0.65 and the sample size is 100, the difference of the fitted *R* squared values between the EWUI model and the MMMUI is 0.01. Table 17 shows that under the condition that the *R* squared values of the fitted models are 0.5 and the sample size is 100, the difference of the fitted *R* squared values between the EWUI model and the MMMUI is 0.055. That explains the reason that the power increases when the *R* squared values of the fitted models decrease under the condition that the sample size is 100.

Table 16 presents the Type I error rate when the generating model is the MMMUI and the comparison model is the FMMUI. Under most conditions, the Type

I error rate is a little higher than 0.05. There is no clear pattern in Table 16. By using the upper bound of 7.5%, the result is acceptable. When the generating model is the MMMUI and the fitted model is the FMMUI, the null hypothesis of the MMMUI as the generating model is retained most of the time. The MMMUI and the PODUI model have the same degrees of freedom. When the MMMUI is compared with the FMMUI, the Type I error rate does not inflate. The results in Table 16 verify the speculation in Table 13 that the when the PODUI is the generating model, the iterative method for fitting the model is related to the inflated Type I error rate in Table 13.

Table 18

The Power When the Generating Model is the FMMUI Under the Condition that the Rank Orders of the Regression Weights Are Not the Same for the Reference Group and the Focal Group

Sample Size	Comparison Model	Generating *R* squared		
		Reference Group 0.2 Focal Group 0.5	Reference Group 0.5 Focal Group 0.8	Reference Group 0.8 Focal Group 0.2
100	EWEI	78.3%	99.9%	99.7%
	EWUI	74.8%	99.9%	99.5%
	PODUI	55.4%	99.5%	73.4%
	MMMUI	62.3%	99.7%	90.3%
500	EWEI	100%	100%	100%
	EWUI	100%	100%	100%
	PODUI	100%	100%	100%
	MMMUI	100%	100%	100%
1000	EWEI	100%	100%	100%
	EWUI	100%	100%	100%
	PODUI	100%	100%	100%
	MMMUI	100%	100%	100%

Table 19

Values of R Squared, Adjusted R Squared, and the F Statistic When the Generating Model is the FMMUI Under the Condition that the Rank Orders of the Regression Weights Are Not the Same for the Reference Group and the Focal Group

Sample Size	Generating *R* Squared Reference	Generating *R* Squared Focal	Fitted Model	Fitted *R* Squared	Fitted Adjusted *R* Squared	*F* Statistic
100	0.2	0.5	EWEI	0.316	0.302	23.03
			EWUI	0.330	0.313	19.50
			PODUI	0.351	0.331	17.73
			MMMUI	0.347	0.326	17.42
			FMMUI	0.381	0.351	13.25
	0.5	0.8	EWEI	0.593	0.584	72.47
			EWUI	0.604	0.594	60.63
			PODUI	0.616	0.604	52.78
			MMMUI	0.614	0.602	52.25
			FMMUI	0.665	0.650	42.82
	0.8	0.2	EWEI	0.424	0.412	36.76
			EWUI	0.437	0.422	30.85
			PODUI	0.492	0.476	31.76
			MMMUI	0.481	0.465	30.44
			FMMUI	0.524	0.502	23.72
500	0.2	0.5	EWEI	0.308	0.305	111.20
			EWUI	0.319	0.315	93.42
			PODUI	0.338	0.334	84.92
			MMMUI	0.334	0.329	83.15
			FMMUI	0.360	0.354	62.19
	0.5	0.8	EWEI	0.589	0.587	357.62
			EWUI	0.599	0.600	298.34
			PODUI	0.609	0.607	259.42
			MMMUI	0.607	0.604	256.58
			FMMUI	0.654	0.651	208.76
	0.8	0.2	EWEI	0.418	0.416	179.71
			EWUI	0.429	0.426	149.92
			PODUI	0.484	0.481	156.03
			MMMUI	0.471	0.468	148.22
			FMMUI	0.509	0.504	114.34

Models of Moderation in Regression

Sample Size	Generating *R* Squared Reference	Generating *R* Squared Focal	Fitted Model	Fitted *R* Squared	Fitted Adjusted *R* Squared	*F* Statistic
1000	0.2	0.5	EWEI	0.307	0.306	221.85
			EWUI	0.318	0.316	186.00
			PODUI	0.337	0.335	169.12
			MMMUI	0.332	0.330	165.55
			FMMUI	0.358	0.355	123.53
	0.5	0.8	EWEI	0.589	0.588	715.71
			EWUI	0.599	0.598	596.71
			PODUI	0.610	0.608	518.97
			MMMUI	0.607	0.605	513.20
			FMMUI	0.654	0.652	417.86
	0.8	0.2	EWEI	0.418	0.416	358.44
			EWUI	0.428	0.426	298.80
			PODUI	0.483	0.482	311.26
			MMMUI	0.470	0.469	295.29
			FMMUI	0.507	0.504	227.42

Table 19 presents the results of the values of the fitted *R* squared, adjusted *R* squared, and other statistics when the generating model is the FMMUI under the condition that the rank orders of the regression weights are not the same for the reference group and the focal group. When the generating *R* squared values are 0.2 for the reference group and 0.5 for the focal group, the *R* squared values for all the fitted models are around 0.35, the average value of 0.2 and 0.5. When the generating *R* squared values are 0.5 for the reference group and 0.8 for the focal group, the *R* squared values for all the fitted models are around 0.65, the average value of 0.5 and 0.8. When the generating *R* squared values are 0.8 for the reference group and 0.2 for the focal group, the *R* squared values for all the fitted models are around 0.5, the average value of 0.8 and 0.2.

Table 18 presents the power when the generating model is the FMMUI under the condition that the rank orders of the regression weights are not the same for the reference group and the focal group. In general, sample size is related to adequate power.

When the sample size is 100 and the comparison model is the EWEI model, the power is lower than 80% if the generating *R* squared values of the reference group are 0.2. Under this condition, Table 19 shows that the *R* squared values for the fitted models are about 0.35. The power is greater than 90% when the *R* squared values of the reference groups are 0.5 and 0.8. Under these conditions, Table 19 shows that the *R* squared values for the fitted models are 0.65 and 0.45.

When the sample size is 100 and the comparison model is the EWUI model, the power is lower than 80% if the generating *R* squared values of the reference group are 0.2. Under this condition, Table 19 shows that the *R* squared values of the fitted models are about 0.35. The power is greater than 90% when the generating *R* squared values of the reference groups are 0.5 and 0.8. Under this condition, Table 19 shows that the *R* squared values for the fitted models are 0.65 and 0.45.

When the comparison model is the EWEI or the EWUI models and the sample size is 100, the lower power is related to the lowest fitted *R* squared value. When the *R* squared values for the fitted models are greater than 0.6, the power increases. When the *R* squared values for the fitted models are greater than 0.4, the power also increases. Table 19 shows that when the fitted *R* squared values are about 0.45, the

difference of the fitted *R* squared values between the EWEI model and the FMMUI is 0.1, and the difference of the *R* squared values between the EWUI model and the FMMUI is 0.087. Then, the power also increases. Thus, two factors, a larger difference of the fitted *R* squared values between the two models compared, and a larger value of the fitted *R* squared for the generating model are related to adequate power.

When the sample size is 500 or 1000 and the comparison model is the PODUI model, under each condition, the power is greater than 99%. When the sample size is 100 and the comparison model is the PODUI model, the power is lower than 80% if the generating *R* squared values of the reference groups are 0.2 and 0.8. Table 19 shows that the *R* squared values of the fitted models are are 0.35 and 0.45 if the generating *R* squared values are 0.2 and 0.8. The power is greater than 90% when the generating *R* squared value of the reference group is 0.5. Table 19 shows that the *R* squared values of the fitted models are 0.65. The lowest power is related to the lowest fitted *R* squared value. When the generating *R* squared value of the reference group is 0.8 and the sample size is 100, Table 19 shows that the difference of the fitted *R* squared values between the PODUI model and the FMMUI is 0.032. Then, the power decreases to 73.4%. Thus, the sufficient *R* squared value is related to adequate power. If the fitted *R* squared values are less than 0.65, the larger difference of the fitted *R* squared value between the two fitted models is related to adequate power.

When the sample size is 500 or 1000 and the comparison model is the MMMUI

model, under each condition, the power is greater than 99%. When the sample size is 100 and the comparison model is the MMMUI, the power is lower than 80% if the generating *R* squared value of the reference group is 0.2. Table 19 shows that under that condition, the *R* squared values of the fitted models are 0.35. The power is greater than 90% when the *R* squared values of the reference groups are 0.5 and 0.8. Table 19 shows that the *R* squared values of the fitted models are 0.65 and 0.45. Table 18 shows that when the comparison model is the MMMUI and the sample size is 100, the lowest power is related to the lowest fitted *R* squared value. When the *R* squared values of the fitted models are greater than 0.6, the power is the highest, about 99%. When the *R* squared values of the fitted models are about 0.45, the power decreases to 90%. Table 19 shows that under this condition, the *R* squared values and the difference of the *R* squared values between the MMMUI and the FMMUI is 0.043. Then, the power is higher than 90% but lower than 99%. Thus, adequate power is related to the sufficient difference of the fitted *R* squared values and the sufficient fitted *R* squared values. Note also the increased power for differentiating FMMUI from the two simplest models EWEI and EWUI (as expected).

Table 20

The Power When the Generating Model is the FMMUI Under the Condition that the Rank Orders of the Regression Weights Are the Same for the Reference Group and the Focal Group

Sample Size	Comparison Model	Generating *R* squared		
		Reference Group 0.2 Focal Group 0.5	Reference Group 0.5 Focal Group 0.8	Reference Group 0.8 Focal Group 0.2
100	EWEI	53.3%	70.5%	96.8%
	EWUI	41.0%	51.0%	95.4%
	PODUI	13.1%	24.9%	20.1%
	MMMUI	27.8%	47.0%	18.4%
500	EWEI	100%	100%	100%
	EWUI	99.7%	100%	100%
	PODUI	35.6%	85.3%	50.5%
	MMMUI	91.6%	99.5%	66.0%
1000	EWEI	100%	100%	100%
	EWUI	100%	100%	100%
	PODUI	57.3%	99.3%	77.3%
	MMMUI	99.6%	100%	95.2%

Table 21

Values of R Squared, Adjusted R Squared, and the F Statistic When the Generating Model is the FMMUI Under the Condition that the Rank Orders of the Regression Weights Are the Same for the Reference Group and the Focal Group

Sample Size	Generating *R* Squared Reference	Generating *R* Squared Focal	Fitted Model	Fitted *R* Squared	Fitted Adjusted *R* Squared	*F* Statistic
100	0.2	0.5	EWEI	0.338	0.324	24.42
			EWUI	0.351	0.334	21.41
			PODUI	0.370	0.350	19.27
			MMMUI	0.364	0.344	18.78
			FMMUI	0.382	0.353	13.33
	0.5	0.8	EWEI	0.634	0.627	86.38
			EWUI	0.646	0.636	72.30
			PODUI	0.655	0.645	62.58
			MMMUI	0.651	0.640	61.30
			FMMUI	0.665	0.650	42.88
	0.8	0.2	EWEI	0.451	0.440	40.99
			EWUI	0.463	0.449	34.29
			PODUI	0.512	0.497	34.40
			MMMUI	0.513	0.498	34.54
			FMMUI	0.524	0.502	23.71
500	0.2	0.5	EWEI	0.328	0.326	122.05
			EWUI	0.338	0.335	102.25
			PODUI	0.356	0.352	91.94
			MMMUI	0.349	0.345	89.01
			FMMUI	0.360	0.354	62.25
	0.5	0.8	EWEI	0.630	0.628	424.30
			EWUI	0.640	0.638	354.52
			PODUI	0.649	0.647	307.02
			MMMUI	0.643	0.641	299.94
			FMMUI	0.654	0.651	209.10
	0.8	0.2	EWEI	0.444	0.442	199.91
			EWUI	0.455	0.452	166.64
			PODUI	0.504	0.501	169.12
			MMMUI	0.503	0.500	168.41
			FMMUI	0.509	0.504	114.38

Models of Moderation in Regression

Sample Size	Generating *R* Squared Reference	Generating *R* Squared Focal	Fitted Model	Fitted *R* Squared	Fitted Adjusted *R* Squared	*F* Statistic
1000	0.2	0.5	EWEI	0.328	0.327	244.11
			EWUI	0.338	0.337	204.31
			PODUI	0.355	0.353	183.43
			MMMUI	0.348	0.346	177.70
			FMMUI	0.358	0.356	123.76
	0.5	0.8	EWEI	0.630	0.630	850.83
			EWUI	0.640	0.639	710.48
			PODUI	0.649	0.648	615.50
			MMMUI	0.644	0.643	601.21
			FMMUI	0.654	0.652	418.58
	0.8	0.2	EWEI	0.443	0.442	397.82
			EWUI	0.453	0.452	331.30
			PODUI	0.503	0.502	337.27
			MMMUI	0.502	0.500	335.38
			FMMUI	0.507	0.504	227.47

Table 21 presents the results of the values of the fitted *R* squared, adjusted *R* squared and other statistics when the generating model is the FMMUI under the condition that the rank orders of the regression weights are the same for the reference group and the focal group. When the generating R-squared values are 0.2 for the reference group and 0.5 for the focal group, the *R* squared values for all the fitted models are around 0.35, the average value of 0.2 and 0.5. When the generating *R* squared values are 0.5 for the reference group and 0.8 for the focal group, the *R* squared values for all the fitted models are around 0.65, the average value of 0.5 and 0.8. When the generating *R* squared values are 0.8 for the reference group and 0.2 for the focal group, the *R* squared values for all the fitted models are around 0.5, the average value of 0.8 and 0.2.

Table 20 presents the power when the generating model is the FMMUI under the condition that the rank orders of the regression weights are the same for the reference group and the focal group. When the comparison models are the EWEI model and the EWUI model, the power is large if the sample size is 500 or 1000. Under the condition that the fitted models are the EWEI and the EWUI models and the sample size is 100, the power increases when the generating *R* squared value of the reference group increases.

Table 20 shows the power when the FMMUI (with the same rank orders of the regression weights for the reference group and the focal group) is the generating model and the alternative hypothesis so the null hypothesis should be rejected. In Table 20, the comparison model is the PODUI model or the MMMUI. When the comparison model is the PODUI model, under most conditions, the power is lower than 80%. Under the condition that the sample size is 1000 and the generating *R* squared value of the reference group is 0.5, the power is the highest, greater than 90%. When the comparison model is the MMMUI, under most conditions, the power is lower than 80%. When the sample size increases to 1000, under each condition, all of the power values are greater than 90%. It is expected that the power is not adequate, because the rank orders of the regression coefficients of the reference group and the focal group for the comparison models (the PODUI model or the MMMUI) are the same as the rank orders for the generating model. Thus, it is more difficult to differentiate the FMMUI model from the simpler models.

Table 22

The Total Number of the Generating Models Identified among 1000 Replications

Sample Size	Generating Model	*R* squared		
		Reference Group 0.2 Focal Group 0.2 or 0.5	Reference Group 0.5 Focal Group 0.5 or 0.8	Reference Group 0.8 Focal Group 0.2 or 0.8
100	EWEI	893	870	886
	EWUI	340	516	798
	PODUI	470	518	843
	MMMUI	549	548	925
	FMMUI	498	994	728
	Different Rank Orders FMMUI Same Rank Orders	102	235	76
500	EWEI	894	903	908
	EWUI	853	864	877
	PODUI	927	916	863
	MMMUI	948	941	933
	FMMUI	1000	1000	1000
	Different Rank Orders FMMUI Same Rank Orders	356	853	293

Models of Moderation in Regression

Sample Size	Generating Model	*R* squared		
		Reference Group 0.2 Focal Group 0.2 or 0.5	Reference Group 0.5 Focal Group 0.5 or 0.8	Reference Group 0.8 Focal Group 0.2 or 0.8
1000	EWEI	905	907	911
	EWUI	896	885	906
	PODUI	918	918	864
	MMMUI	942	941	952
	FMMUI	1000	1000	1000
	Different Ranks Orders FMMUI Same Rank Orders	573	993	726

Note. For Reference Group 0.2 and Focal Group 0.2 or 0.5, Focal group R^2 is 0.2 for the EWEI and EWUI and 0.5 for all other models. For Reference Group 0.5 and Focal group 0.5 or 0.8, Focal group R^2 is 0.5 for the EWEI and EWUI and 0.8 for all other models. For Reference Group 0.8 and Focal Group 0.2 or 0.8, Focal group R^2 is 0.8 for the EWEI and EWUI and 0.2 for all other models.

Table 22 shows the number of replications in which the generating model was identified after comparison to all other models. Table 22 is used to answer the third research question: If the hypothesis is a model, how well can we identify the generating model when the *F* statistic is used to compare several models? To answer this question, I compare each generating model with other models simultaneously. The data set is the one described in the methods section and used for Table 7 to Table 21.

As described above, the generating model is said to be “identified” in a

replication, if two conditions hold in the replication. First, the null hypothesis of the generating model is retained when compared to every other model that includes the generating model as a special case. Second, the null hypothesis is rejected when the generating model (the alternative hypothesis) is compared to every model that is a special case of the generating model. Somewhat arbitrarily, identification is said to be adequate if the generating model was identified 80% of the time, i.e., in at least 800 of 1000 replications.

When the generating model is the EWEI model, the larger fitted models, the EWUI model, the PODUI model, the MMMUI and the FMMUI should not fit significantly better than the EWEI model. If the EWEI model is the generating model, the EWEI hypothesis should not be rejected when it is compared with larger models.

When the generating model is the EWEI model, under all of the conditions, among 1000 replications, the generating model was identified in more than 800 replications. When the sample size increases to 1000, among 1000 replications, the generating model was identified in more than 900 replications.

When the generating model is the EWUI model, the EWUI model should fit significantly better than the EWEI model, and the larger models, the PODUI model, the MMMUI, and the FMMUI should not fit significantly better than the EWUI model. If the EWUI model is the generating model, the EWEI hypothesis should be rejected when it is compared with the EWUI model, and the EWUI hypothesis should not be rejected when it is compared with the PODUI model, the MMMUI model, and

the FMMUI model.

When the generating model is the EWUI model, under the condition that the sample size is 100, among 1000 replications, the generating model was identified in less than 800 replications. When the *R* squared value of the reference group is 0.2, among 1000 replications, the generating model was identified in only 340 replications. When the *R* squared value of the reference group is 0.5, among 1000 replications, the generating model was identified in 548 replications. When the sample size increases to 500 and 1000, among 1000 replications, the generating model was identified in more than 800 replications.

When the generating model is the PODUI model, the PODUI model should fit significantly better than the EWEI model and the EWUI model, but the PODUI model should not fit significantly better than the FMMUI model. If the PODUI model is the generating model, the hypotheses of the EWEI model and the EWUI model should be rejected when they are compared with the PODUI model, and the PODUI hypothesis should not be rejected when it is compared with the FMMUI model.

When the generating model is the PODUI model, under the condition that the sample sizes is 100 and the *R* squared value for the reference group is 0.2 or 0.5, among 1000 replications, the generating model was identified in less than 600 replications. Under other conditions, among 1000 replications, the generating model was identified in more than 800 replications. The larger sample sizes and the larger difference of *R* squared values between the reference group and the focal group are

related to the larger number of replications in Table 22.

When the generating model is the MMMUI, the MMMUI should fit significantly better than the EWEI model and the EWUI model, but the FMMUI should not fit significantly better than the MMMUI. If the MMMUI model is the generating model, the hypotheses of the EWEI model and the EWUI model should be rejected when they are compared with the MMMUI model, and the MMMUI hypothesis should not be rejected when it is compared with the FMMUI model.

When the generating model is the MMMUI, under the condition that the sample size is 100 and the *R* squared value for the reference group is 0.2 or 0.5, among 1000 replications, the generating model was identified in less than 600 replications. Under other conditions, among 1000 replications, the generating model was identified in more than 800 replications. The larger sample sizes and the larger difference of *R* squared values between the reference group and the focal group are related to the larger number of replications in Table 22.

When the generating model is the FMMUI, the FMMUI should fit significantly better than all of the smaller fitted models, the EWEI model, the EWUI model, the PODUI model, and the MMMUI. If the FMMUI model is the generating model, the hypotheses of the EWEI model, the EWUI model, the PODUI model and the MMMUI model should be rejected when they are compared with the FMMUI model.

When the generating model is the FMMUI and the rank orders of the regression weights are not the same for the reference group and the focal group, under the

condition that the sample size is 100 and the *R* squared value for the reference group is 0.2, among 1000 replications, the generating model was identified in less than 500 replications. Under the condition that the sample size is 100 and the *R* squared value for the reference group is 0.8, the generating model was identified in less than 800 replications. Under the condition that the sample size was 100 and the *R* squared value for the reference group is 0.5, the generating model was identified in more than 900 replications. The explanation is: the *R* squared values of fitted models are 0.3, 0.65, and 0.5 respectively, the average values of the generating *R* squared values for the reference group and the focal group under each condition in Table 22. The fitted *R* squared values will influence the identification of the FMMUI model. When the sample size increases to 500 or 1000, under each condition, among 1000 replications, the generating model was identified in 1000 replications.

When the generating model is the FMMUI and the coefficients have the same rank order in both the focal and reference groups, under most conditions, the generating model cannot be well identified (Table 22). Under the condition that the sample size is 500 and the *R* squared for the reference group is 0.5, among 1000 replications, the generating model was identified in more than 800 replications. Under the condition that the sample size is 1000 and the *R* squared for the reference group is 0.5, among 1000 replications, the generating model was identified in more than 900 replications. Under these two conditions, the *R* squared values of fitted models are 0.65, higher than the fitted *R* squared values of the conditions in the other two

columns in Table 22. This result indicates that the generating model can be identified when the fitted *R* squared value is higher and the sample size is large enough. This result also indicates that when the regression weights of the generating model for the reference group and the focal group have the same rank order, the FMMUI as the generating model cannot be identified under most conditions. For example, when the sample size is 100 and the *R* squared for the reference group is 0.2, among 1000 replications, the generating model was identified in approximately 100 replications.

The principles of identification of one of the models used in Table 22 can be applied to Park's (2018) study. The author had two sets of analyses and compared models with others based on the nested structure. As stated in the literature review, Table 23 and Table 24 summarized the results of Park's real data study of model comparisons using the models in my simulation. Table 23 is based on regression models that do not include prior achievement as a covariate. Table 24 is based on regression models that include prior achievement as a covariate.

Table 23 and Table 24 summarizes the results of Park's study of model comparisons of the models used in the simulation study.

Table 23

The Result of Model Comparisons without Prior Achievement as a Covariate

Model	EWEI	EWUI	PODUI	MMMUI	FMMUI
EWEI	NA	F=49.352	NA	NA	NA
		P<0.001			
EWUI	F=49.352	NA	F=1.626	F=2.064	F=0.984
	P<0.001		P<0.001	P<0.001	P=0.693
PODUI	NA	F=1.626	NA	NA	F=0.931
		P<0.001			P=0.987
MMMUI	NA	F=2.064	NA	NA	F=0.894
		P<0.001			P=1.000
FMMUI	NA	F=0.984	F=0.931	F=0.894	NA
		P=0.693	P=0.987	P=1.000	

Note. The author did not compare the models with equal intercept and unequal intercept. The author did not compare the PODUI and MMMUI because they are not hierarchically arranged. NA means that the author did not provide the statistic.

In Table 23,[1] the EWUI model was statistically significant over the EWEI model. The PODUI model was statistically significant over the EWUI model. The MMMUI model was statistically significant over the EWUI model. The FMMMUI is not statistically significant when compared to the PODUI model or the MMMUI. Thus, by using the principle in Table 22, the PODUI model or the MMMUI would be identified as the "optimal" model.

[1] This table has the results from Park's (2018) dissertation.

Table 24

The Result of Model Comparison with Prior Achievement As the Covariate

Model	EWEI	EWUI	PODUI	MMMUI	FMMUI
EWEI	NA	F=10.582 P<0.001	NA	NA	NA
EWUI	F=10.582 P<0.001	NA	F=0.434 P=1.000	F=0.315 P=1.000	F=0.684 P=1.000
PODUI	NA	F=0.434 P=1.000	NA	NA	F=0.704 P=1.000
MMMUI	NA	F=0.315 P=1.000	NA	NA	F=0.713 P=1.000
FMMUI	NA	F=0.684 P=1.000	F=0.704 P=1.000	F=0.713 P=1.000	NA

Note. The author did not compare the models with equal intercept and unequal intercept. The author did not compare the PODUI and MMMUI because they are not hierarchically arranged. NA means that the author did not provide the statistic.

In Table 24,[2] the EWUI model was statistically significant when compared to the EWEI model. None of the three models, the PODUI model, the MMMUI and the FMMUI was statistically significant when compared to the EWUI model. Thus, by using the principle provided by Table 22, the EWUI model is the optimal model.

Discussion

The three research questions in these study are: 1) When two models are compared using the *F* statistic, what will be the Type I error rate when the generating model is the less general model of the two fitted models? 2) When two models are

[2] This table has the results from Park's (2018) dissertation.

compared with the *F* statistic, what will be the power when the generating model is the more general of the two fitted models? 3) How well can we identify the generating model when the *F* statistic is used to compare the several models?

To answer the three questions, a simulation study with five generating models was conducted. The sample sizes (100, 500, and 1000) and the generating *R* squared values for the reference and focal group were varied.

In the result part, Table 7, Table 10, Table 13, and Table 16 answered the first question. Table 9, Table 12, Table 15, Table 18, and Table 20 answer the second question. Table 22 answers the third question.

This chapter will cover, 1) the influence of the sample size on the Type I error rate, power and identification of the hypothesized generating model, 2) the influence of the generating model *R* squared values of the reference group and the focal group on the Type I error rate, power and identification of the hypothesized generating model, 3) the influence of the difference of the generating *R* squared values between the reference group and the focal group on the Type I error rate, power, and identification of the hypothesized generating model and, 4) the strength, limitation, and suggestions from this study.

The influence of the sample size

By using 80% or above as the standard for power, among generating models, power was greatest for the EWUI model under most conditions. Only under the condition that the sample size is 100 and the *R* squared values of the generating model

are 0.2 and 0.5, was the power lower than 80%.

The PODUI model, MMMUI, and the FMMUI are the models with different regression weights for the reference group and the focal group. When the generating model is the PODUI model and the sample size is large enough (500 or 1000), in Table 12, the power is greater than 80%. When the generating model is the MMMUI, in Table 15, when the sample size is large enough (500 or 1000), the power is large enough. When the generating model is the FMMUI with different rank orders of regression weights for the reference group and the focal group, in Table 18, if the sample size is large enough (500 or 1000), for all the comparison models, the power is sufficient.

For each generating model, under each condition, Table 22 provides evidence of the sufficient sample size to identify the model as the hypothesis when compared to all other models. When the model as the hypothesis is the EWEI model, the sample size 100 is sufficient to identify the generating model. When the model as the hypothesis is the EWUI model, the PODUI, the MMMUI, or the FMMUI with different rank orders, the sample size 500 or 1000 is always sufficient to identify the model as the hypothesis. When the generating model is the FMMUI with the same rank order of regression weights for the reference group and the focal group, the generating FMMUI model can be identified only under the conditions that the sample size is 500 or 1000 and the *R* squared value for the reference and focal groups are 0.5 and 0.8 respectively.

The influence of the *R* squared values for the reference and the focal group

For the generating model with the same regression weights for the reference group and the focal group, the EWEI model and the EWUI model, the *R* squared value for the generating model will be the same for the reference group and the focal group. Then, the fitted *R* squared value will be close to the *R* squared value for the reference group and the focal group for the generating model. Though the sample size 100 is related to lower power when the generating model is the EWUI model in Table 11, the *R* squared value of the reference group is related to adequate power. When the sample size is 100, in Table 22, if the models as the hypothesis are the EWEI model and the EWUI model, the models with the same regression weights for the reference group and the focal group, the larger number of identifications among 1000 replications is associated with the larger *R* squared value of the reference group (0.5 or 0.8). Thus, if the sample size is 100, the *R* squared value 0.8 of the reference group is sufficient to identify the model as the hypothesis.

For the generating model with different regression weights for the reference group and the focal group, the PODUI model, the MMMUI, and the FMMUI, the fitted *R* squared value will be the average value of the two values for the generating model(s). When the sample size is 100, if the models as the hypothesis are the PODUI model, the MMMUI, and the FMMUI with different rank orders for the reference group and the focal group, the models with the different regression weights for the reference group and the focal group, in Table 22, the larger number of identifications

is associated with the *R* squared value 0.8 for the reference group and 0.2 for the focal group. The value 0.8 for the reference group is sufficient for the identification of the model as the hypothesis with the sample size 100 under most conditions. One exception is the FMMUI with the same rank order of regression weights for the reference group and the focal group.

When the generating model is the FMMUI with the same rank order of regression weights for the reference group and the focal group, in Table 20, for the comparison models with the same regression weighs for the reference group and the focal group, the EWEI model and the EWUI model, the larger sample size (500 or 1000) is associated with adequate power. Under each condition, the comparison models with different regression weights for the reference group and the focal group, the PODUI model and the MMMUI, cannot be differentiated from the generating model. Even under the condition that the sample size is 1000, the power is still lower than 60% when the *R* squared value for the reference group is 0.2 and the comparison model is the PODUI model.

In general, under each condition, the lowest fitted *R* squared value is associated with the lowest power in each Table. However, the highest fitted *R* squared value may not be associated with the highest power. The larger difference of *R* squared values between the reference group and the focal group may be associated with the larger power.

The influence of the difference of the generating *R* squared values between the reference group and the focal group

For each generating model with different regression weights for the reference group and the focal group, the larger difference of the *R* squared values between the reference group and the focal group for the generating model is related to the fitted *R* squared difference between the generating model and the comparison model.

In Table 12 and Table 15, when the generating model is the PODUI model or the MMMUI model and the fitted model is the EWEI model or EWUI model, sample size 100 is sufficient to yield adequate power if the *R* squared value for the reference group is 0.8 and 0.2 for the focal group. Though the fitted *R* squared values are about 0.45, the average value of 0.8 and 0.2 the difference between the fitted *R* squared values of the generating model and the fitted model is greater than 0.05. Thus, the power increases.

When the generating model is the FMMUI, under the condition that the rank orders of the regression weights are not the same for the reference group and the focal group, both the fitted *R* squared value and the difference of the *R* squared values between the reference group and the focal group influences the power. If the fitted model is the EWEI model or the EWUI model, the larger difference of the *R* squared values between the generating model and the comparison model increases the power. If the fitted model is the PODUI model or the MMMUI, the higher fitted *R* squared value is related to the higher power.

When the generating model is the FMMUI, under the condition that the rank orders of the regression weights are the same for the reference group and the focal group, the trend of the power is not very obvious. I will talk about it in the conclusion part.

Strength, limitation, and suggestions

The strength of this study is to provide a guide of model selection under the condition that the two models are compared and one of the models is compared with all the other models. That was not provided in Park's (2018) empirical study. Additionally, the method of generating models was not presented previously. The method used in this study to generate the PODUI and the MMMUI provided an example for other researchers.

In this study, a few results should be investigated in the future because their explanation is not clear.

When the data was generated, to make the paper manageable, the sample sizes for each group were set to be equal. In future research, the sample sizes for the reference group and the focal group should vary.

For each generating model, to make this study manageable, the inter-correlations among predictors were fixed. Though the inter-correlations 0.3 may not cause a problem of multicolinearity, we do not know if higher correlations between predictors can impact the result. In future research, the inter-correlations among predictors should be varied.

In this study, four continuous predictors and one categorical predictor were used in the analysis. In future studies, the number of predictors should be varied. For example, the number of categorical predictors can be increased, and the number of focal groups for each categorical predictor also can be increased.

When the generating model was the EWEI model or the EWUI model, Table 8 and Table 11 showed that under the condition that the generating *R* squared value was 0.5, the fitted *R* squared values were 0.6, higher than the expected value 0.5. In other tables with 0.5 as the generating *R* squared value, this problem did not exist. The reason for the higher fitted *R* squared values in Table 8 and Table 11 should be explored in future studies.

When the generating model is the EWUI model and the comparison model is the PODUI model or the MMMUI, under each condition, the Type I error rate inflated to 6% to 7%, a little higher than the expected value 5%. One possible reason is when the comparison model has only one more parameter than the generating model, but the generating model is a special case of the comparison model, the Type I error rate for the comparison model may be higher. Additional research will be required to explain this result.

When the generating model is the FMMUI under the condition that the rank orders of the regression weights are the same for the reference group and the focal group, the generating model is not powerful under most conditions. When the comparison model is the PODUI or the MMMUI, under the condition that the *R*

squared value of the reference and focal groups are 0.8 and .2 respectively, with the fitted *R* squared value about 0.45, the power can be lower than 80% even though the sample size is 500. Possible reasons are the fitted *R* squared values and the difference of the fitted *R* squared values between the generating model and the comparison model, which can affect the power. Under certain condition, to get adequate power, the difference of the fitted *R* squared values between the generating model and the comparison model should attain a threshold. That speculation will also need additional analyses in the future.

Conclusion

In this study, there were two independent variables: the sample size and *R* squared for the reference and focal group. There were two variables, the fitted R squared value of the generating model and the fitted *R* squared values between the two models compared, that seem to mediate the effect of the independent variables. The *F* statistic in the literature review section can also be written as

$$F = \frac{R_G^2 - R_S^2}{1 - R_G^2} * \frac{DF_G}{DF_G - DF_S} \quad (15)$$

For the *R* squared, the subscript G represents the larger model, and the subscript *S* represents the smaller model. For the degrees of freedom, the *DF* represented the degrees of freedom of the model.

The first mediating variable is the fitted *R* squared for the more general of the two models being compared, which is the factor in the denominator of the *F* statistic in equation 15. The larger this fitted *R* squared, the greater the power and the better

the identification of the correct model when the more general of the two models was the generating model. The second mediating factor is the difference in the fitted *R* squared for the two models being compared, which is the numerator of the *F* statistic in equation 15. The larger the difference, the greater the power and the better the identification of the correct model when the more general model was the correct model. These two factors seemed to mediate the effects of the independent variables, and the mediation effects are readily explained by the numerator and denominator of equation 15. For the generating models with different regression weights for the reference group and the focal group, the higher fitted *R* squared value is related to the identification of the generating models. For the generating models with different regression weights for the reference group and the focal group, the difference of *R* squared values between the reference group and the focal group is related to the difference of the fitted *R* squared values between the generating model and the comparison model. That also influences the identification of the generating models in Table 22.

In the discussion section, the inflated Type I error rate when the generating model was the EWUI and the comparison models was the PODUI or MMMUI was presented. One possibility is when the comparison model has only one more parameter than the generating model, but the generating model is a special case of the comparison model, the Type I error rate for the comparison model can be inflated slightly. For the trend of the power when the generating model was the FMMUI under

the condition that the rank orders of the regression weights are the same for the reference group and the focal group, either the fitted *R* squared value or the difference of the fitted *R* squared values between the generating model and the comparison model could affect the power. Under certain conditions, to get adequate power, the difference of the fitted *R* squared values should attain a threshold. That speculation will also need additional analyses in the future.

Only under the condition that the generating model is the EWEI model, may the sample size 100 be sufficient. The major finding is that under most conditions, a sample size 500 or above is needed to identify the generating models in at least 80% of replications.

References

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

Aguinis, H., Pierce, C. A., & Stone-Romero, E. F. (1994). Estimating the power to detect dichotomous moderators with moderated multiple regression. *Educational and Psychological Measurement,* 54, 690-692.

Aguinis, H., & Stone-Romero, E. F. (1997). Methodological artifacts in moderated multiple regression and their effects on statistical power. *Journal of Applied Psychology*, *82*(1), 192–206.

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. SAGE publications, Inc.

Blanca, M. J., Alarcón, R., Arnau, J., Bono, R., & Bendayan, R. (2018). Effect of variance ratio on ANOVA robustness: Might 1.5 be the limit? *Behavior Research Methods, 50*(3), 937–962.

Busemeyer, J. R., & Jones, L. E. (1983). Analysis of multiplicative combination rules when the causal variables are measured with error. *Psychological Bulletin, 93*(3), 549–562.

Cook, R. D., & Weisberg, S. (2004). Partial One-Dimensional Regression Models. *The American Statistician*, *58*(2), 110–116. https://doi.org/10.1198/0003130043240

Davison, M. L., & Davenport, E. C. (2002). Identifying criterion-related patterns of

predictor scores using multiple regression. *Psychological Methods, 7(4)*, 468–484. https://doi.org/10.1037/1082-989X.7.4.468

Davison, M. L., Davenport, E. C., & Kohli, N. (2017). Population Differences in Linear Regression Coefficients: The Mean Moderated Model, Personal Communication.

Davison, M. L., Davenport, E. C., & Hao, J. (2022). Linear Equality Constraints: Reformulations of Criterion Related Profile Analysis with Extensions to Moderated Regression for Multiple Groups. *Psychological Methods.* https://doi.org/10.1037/met0000430.

Davison, M. L., Davenport, E. C., Jr., Kohli, N., Kang,Y., & Park, K. (2020). Addressing quantitative and qualitative hypotheses using regression models with equality restrictions and predictors measured in common units. *Multivariate Behavioral Research.* https://doi.org/10.1080/00273171.2020.1754154

Davison, M. L., Jew, G., & Davenport, E. C., Jr. (2014a). Patterns of SAT Scores, Choice of STEM Major, and Gender. *Measurement and Evaluation in Counseling and Development, 47,* 118 – 126. https://doi.org/10.1177/0748175614522269

Dunlap, W. P., & Kemery, E. R. (1988). Effects of predictor intercorrelations and reliabilities on moderated multiple regression. *Organizational Behavior and Human Decision Processes, 41*(2), 248–258.

Evans, M. G. (1985). A Monte Carlo study of the effects of correlated method variance in moderated multiple regression analysis. *Organizational Behavior and Human Decision Processes, 36*(3), 305–323.

Kang, Y. (2022). *Models of Moderation in Regression.* Unpublished Master Thesis. University of Minnesota

McClelland, G. H., & Judd, C. M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin, 114*(2), 376–390.

Park, K. (2018). *Using Multiple Regression to Ascertain Group Differences in the Relationship of Predictors to a Criterion: Ethnic Group Differences in the Relationship between Course-taking and Achievement in Mathematics.* Unpublished Doctoral Dissertation. University of Minnesota

Paunonen, S. V., & Jackson, D. N. (1988). Type I error rates for moderated multiple regression analysis. *Journal of Applied Psychology, 73,* 569-573.

Saunders, D. R. (1956). Moderator variables in prediction. *Educational and Psychological Measurement*, *16*(2), 209–222.

Stone-Romero, E. F., & Anderson, L. E. (1994). Relative power of moderated multiple regression and the comparison of subgroup correlation coefficients for detecting moderating effects. *Journal of Applied Psychology, 79*(3), 354–359.

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*(1), 167–178.

Wiernik, B. M., Wilmot, M. P., Davison, M. L., & Ones, D. S. (2021). Meta-analytic criterion profile analysis. *Psychological Methods, 26(2),* 186–209. https://doi.org/10.1037/met0000305