ChinaRxiv 中X

AI translation · View original & related papers at
chinarxiv.org/items/chinaxiv-202112.00048

Moderation Analysis and Effect Size Based on Two-Level Regression Models

Authors: Fang Jie, Wen Zhonglin, Wen Zhonglin

Date: 2021-12-09T00:00:00+00:00

Abstract

The use of multiple regression for moderation effect analysis has been frequently applied in the social sciences. This paper briefly describes the current limitations of moderation effect analysis using multiple regression, including artificial transformation of test models, insufficient distinction between independent and moderator variables, difficulty in meeting the assumption of homogeneity of error variance, and the fact that the moderation effect size indicator ΔR^2 does not directly measure the degree to which the moderator variable moderates the relationship between the independent and dependent variables. A better approach is to use two-level regression models for moderation effect analysis and employ corresponding effect size indicators. After introducing the new method and new effect size, a systematic analysis procedure for moderation effects is summarized, and an example is provided to demonstrate how to conduct moderation effect and effect size analysis for two-level regression models using Mplus software. Finally, the development of moderation effect analysis for two-level regression models is discussed, including robust moderation effect analysis, latent variable moderation effect analysis, moderated mediation effect analysis, and mediated moderation effect analysis.

Full Text

Moderation Analysis and Its Effect Size Based on a Two-Level Regression Model

FANG Jie[1], **WEN Zhonglin**[2]
([1]Institute of Guangdong Economy and Social Development & Department of Applied Psychology, Guangdong University of Finance & Economics, Guangzhou 510320)
([2]Center for Studies of Psychological Application & School of Psychology, South China Normal University, Guangzhou 510631)

ChinaRxiv

Abstract
Multiple regression has been frequently used for moderation analysis in social sciences. This paper briefly discusses the limitations of current multiple regression approaches to moderation analysis, including the artificial transformation of the model, insufficient distinction between independent and moderator variables, difficulty meeting the homogeneity of error variance assumption, and the fact that the moderation effect size index ΔR^2 does not directly measure the degree to which the moderator influences the relationship between the independent and dependent variables. A better approach is to conduct moderation analysis using a two-level regression model with corresponding effect size indices. After introducing the new method and effect sizes, we propose a comprehensive analysis procedure for moderation effects. An example demonstrates how to use Mplus software to conduct moderation analysis and compute effect sizes based on a two-level regression model. Finally, we discuss future developments in two-level regression model moderation analysis, including robust moderation analysis, moderation analysis with latent variables, moderated mediation analysis, and mediated moderation analysis.

Keywords: moderating effect, two-level regression model, multiple regression, effect size

Moderator variables play an important role in psychological, educational, social, and management research. If the relationship between independent variable X and dependent variable Y is influenced by a third variable Z, then Z is a moderator. In this case, the relationship between X and Y is a function of the moderator Z, which can be represented by the moderation model shown in Figure 1 Figure 1: see original paper. Multiple regression is currently the most common method for moderation analysis (Wen et al., 2005; Fang et al., 2015), but it has several shortcomings. Yuan et al. (2014) proposed using a two-level regression model for moderation analysis and introduced new effect size indices. This paper provides an in-depth discussion of moderation analysis and its effect sizes based on the two-level regression model. After briefly introducing multiple regression moderation analysis and its effect size index ΔR^2, we discuss the limitations of multiple regression and ΔR^2. We then detail how to conduct moderation analysis using a two-level regression model and introduce new effect size indices corresponding to this model, followed by a proposed analysis procedure. An example demonstrates how to use Mplus software for moderation analysis and effect size computation using a two-level regression model. Finally, we discuss future developments in this area.

Received: 2021-07-29
*This research was supported by the National Natural Science Foundation of China (32171091) and the National Social Science Foundation of China (17BTJ035).

Corresponding Author: Wen Zhonglin, E-mail: wenzl@scnu.edu.cn

(a) Conceptual Diagram (b) Path Diagram

Figure 1. Moderation Model Diagram (adapted from Fang et al., 2015)

2. Multiple Regression Approach to Moderation Analysis and Its Traditional Effect Size

Common moderation effects can be analyzed using the following regression equation (Figure 1(b) shows the corresponding path diagram):

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ + e$$

where β_0 is the intercept, β_1 is the effect of independent variable X on dependent variable Y when Z = 0, β_2 is the effect of moderator Z on Y when X = 0, and the error term e is assumed to follow a normal distribution. If the estimated coefficient β_3 is statistically significant, it indicates that moderator Z moderates the relationship between X and Y. The moderation effect size is measured by:

$$\Delta R^2 = R_2^2 - R_1^2$$

where R_2^2 represents the coefficient of determination for regression equation (1), and R_1^2 represents the coefficient of determination for regression equation (1) after removing the product term XZ (Wen et al., 2005, 2020; Fang et al., 2015).

Currently, multiple regression is widely used for moderation analysis. However, this method has several limitations (Yuan et al., 2014). First, multiple regression artificially transforms the moderation model in Figure 1(a) into an interaction model in Figure 1(b) and estimates this model using ordinary least squares. This approach actually tests whether the interaction between Z and X is significant, rather than directly testing the moderation model in Figure 1(a). Second, the roles of independent and moderator variables are difficult to distinguish in equation (1). Using multiple regression, moderation is symmetric: if the effect of X on Y is moderated by Z (where X is the independent variable), then the effect of Z on Y is also moderated by X (where X becomes the moderator), and both analyses yield identical results (Hayes & Montoya, 2017). Third, the homogeneity of error variance assumption is often violated. Multiple regression assumes homoscedasticity (i.e., error variances are equal across different values of the independent variable X), and significance tests for regression coefficients are based on this assumption. However, equation (1) rarely satisfies this assumption because when X takes a specific value x_i, the error variance in equation (1) becomes:

$$Var(e|X = x_i) = \sigma_e^2 + \beta_3^2 Var(Z|X = x_i)$$

Only when $\beta_3 = 0$ are the error variances equal (independent of x_i) (Liu et al., 2021). Aguinis et al. (1999) found that approximately 50% of articles using multiple regression analysis clearly violate the homogeneity of error variance assumption, which leads to biased standard errors for parameter estimates of β_3, resulting in reduced statistical power for moderation analysis and biased effect size estimates (Liu et al., 2021). Fourth, the moderation effect size index ΔR^2 has limitations. ΔR^2 does not directly measure the degree to which moderator Z influences the X→Y relationship; rather, it reflects the additional explanatory power of the product term XZ on the variance of Y (Liu & Yuan, in press). Wen and Ye (2014) suggested that a ΔR^2 change exceeding 2% (or even 3%) is substantively meaningful, but ΔR^2 values are often small (Aguinis et al., 2005; Liu & Yuan, in press).

3.1. Moderation Analysis Using a Two-Level Regression Model

Moderation analysis based on a two-level regression model (Figure 2 [Figure 2: see original paper]) can be expressed as (Yuan et al., 2014):

Level 1:

$$Y_i = \beta_{0i} + \beta_{1i} X_i + e_i$$

Level 2:

$$\beta_{0i} = \gamma_{00} + \gamma_{01} Z_i + \varepsilon_{0i}$$

$$\beta_{1i} = \gamma_{10} + \gamma_{11} Z_i + \varepsilon_{1i}$$

where β_{0i} and β_{1i} are random coefficients that vary across individuals i, while γ_{00}, γ_{01}, γ_{10}, and γ_{11} are fixed coefficients that do not vary across individuals. e_i, ε_{0i}, and ε_{1i} all follow normal distributions with mean 0. Specifically, $e_i \sim N(0, \sigma_e^2)$, and:

$$\begin{pmatrix} \varepsilon_{0i} \\ \varepsilon_{1i} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{pmatrix} \right)$$

The Level 1 error e_i and Level 2 errors (ε_{0i} and ε_{1i}) are independent, but ε_{0i} and ε_{1i} may be correlated (the covariance of errors is denoted by σ_{01}). The independent variable X_i and moderator Z_i may also be correlated (indicated by a dashed line in Figure 2). Substituting equations (3a) and (3b) into equation (2) yields:

$$Y_i = \gamma_{00} + \gamma_{01} Z_i + \gamma_{10} X_i + \gamma_{11} X_{iZ} i + (\varepsilon_{0i} + \varepsilon_{1i} X_i + e_i)$$

If the coefficient of the product term $X_{iZ} i$, γ_{11}, is significantly different from 0, it indicates a significant moderating effect of Z_i on the relationship between X_i and Y_i. It should be noted that because equations (2)-(4) all have only subscript i, the two-level regression model actually conducts moderation analysis on single-level data (individual level). Because the actual data are not nested two-level data, the two-level regression model cannot distinguish ε_{1i} from e_i, nor can it distinguish β_{0i} from β_{1i}. Therefore, we set $\sigma_1^2 = 0$ and $\sigma_{01} = 0$.

Using a two-level structural model to analyze single-level data for moderation effects offers clear advantages. First, in the two-level regression model (equations (2)-(4), Figure 2), the moderator Z_i directly explains the relationship between the independent variable X_i and dependent variable Y_i (i.e., β_{0i} and β_{1i}), addressing the first limitation of multiple regression. Second, the two-level regression model clearly distinguishes the roles of independent variable X_i and moderator Z_i: X_i appears in the Level 1 equation and explains variance in the dependent variable Y_i, while Z_i appears only in the Level 2 equations and explains variance in the Level 1 intercept β_{0i} and slope β_{1i}. The independent and moderator variables are not interchangeable; they serve distinct functions, addressing the second limitation of multiple regression. Third, the two-level regression model does not require the homogeneity of error variance assumption and can conduct moderation analysis with heteroscedastic error variances. This is because the error variance in the two-level regression model (equation (4)) varies with the independent variable X_i:

$$Var(\varepsilon_{0i} + \varepsilon_{1i}X_i + e_i) = \sigma_0^2 + \sigma_1^2 X_i^2 + \sigma_e^2 + 2\sigma_{01}X_i$$

The two-level regression model inherently accommodates heteroscedastic error variances (Yuan et al., 2014). Simulation studies by Yuan et al. (2014) found that when the homogeneity of error variance assumption is violated (i.e., error variances are heteroscedastic), the two-level regression model provides more accurate parameter estimates than multiple regression. Additionally, when the homogeneity assumption holds, parameter estimates from the two-level regression model are essentially equivalent to those from multiple regression. This is because when $\sigma_1^2 = 0$, the multiple regression moderation model (equation (1)) can be viewed as a special case of the two-level regression model (equation (4)) (Liu & Yuan, in press).

3.2.1. Effect Size Indicator: $R_{\beta 1}^2$

Yuan et al. (2014) proposed that the effect size for moderation in the two-level regression model is:

$$R_{\beta 1}^2 = \frac{\hat{\gamma}_{11}^2 Var(Z_i)}{\hat{\gamma}_{11}^2 Var(Z_i) + \hat{\sigma}_1^2}$$

where $Var(Z_i)$ is the sample variance of moderator Z_i, and $\hat{\gamma}_{11}$ and $\hat{\sigma}_1^2$ are estimated values. $R_{\beta 1}^2$ represents the proportion of variance in β_{1i} that can be explained by moderator Z_i. Since β_{1i} is the regression coefficient of dependent variable Y_i on independent variable X_i, $R_{\beta 1}^2$ represents the degree to which moderator Z_i moderates the relationship between X_i and Y_i. Therefore, using $R_{\beta 1}^2$ as an effect size indicator better aligns with the conceptual definition of moderation in Figure 1(a). In most cases, $R^2_{\beta 1}$ will be larger than the moderation effect size ΔR^2 obtained from multiple regression because $R^2_{\beta 1}$ indicates how much variance in β_{1i} is explained, whereas ΔR^2 indicates how much variance in dependent variable Y is explained, and the variance of β_{1i} is generally smaller than the variance of Y (Liu & Yuan, in press). Moreover, $R^2_{\beta 1}$ ranges from [0, 1]. Specifically, $R^2_{\beta 1} = 1$ indicates that changes in β_{1i} are completely caused by changes in Z_i (complete moderation); $0 < R^2_{\beta 1} < 1$ indicates that changes in β_{1i} are partially caused by changes in Z_i (partial moderation); and $R^2_{\beta 1} = 0$ indicates that Z_i has no moderating effect (Yang & Yuan, 2016). In summary, $R^2_{\beta 1}$ is a more appropriate effect size indicator than ΔR^2, addressing the fourth limitation of multiple regression.

3.2.2. Effect Size Indicators $R^2_{mo\Delta}$ and V^2_R

Liu and Yuan (in press) noted that $R^2_{\beta 1}$ still has limitations. $R^2_{\beta 1}$ represents the proportion of variance in β_{1i} explained by moderator Z_i, rather than the proportion of variance in Y explained by X_i (i.e., $Var(\beta_{1i}X_i)$) that can be explained by Z_i. Therefore, they proposed several new moderation effect size indicators.

The first is the $R^2_{mo\Delta}$ effect size indicator:

$$R^2_{mo\Delta} = \frac{R^2_{21} - R^2_{20}}{1 - R^2_{22}}$$

where R^2_{20} represents the coefficient of determination for equation $Y_i = \gamma_{00} + \gamma_{01}Z_i + e_i$, indicating the proportion of variance in Y_i explained by Z_i. R^2_{21} represents the coefficient of determination for equation $Y_i = \gamma_{00} + \gamma_{01}Z_i + \gamma_{10}X_i + \gamma_{11}X_{iZ}i + e_i$, indicating the proportion of variance in Y_i explained jointly by Z_i, X_i, and $X_{iZ}i$. R^2_{22} represents the coefficient of determination for equation (4), indicating the proportion of variance in Y_i explained by Z_i, X_i, and $X_{iZ}i$.

$R^2_{mo\Delta}$ represents the proportion of variance in Y_i uniquely explained by $X_{iZ}i$ (excluding the effects of X_i and Z_i) relative to the variance in Y_i jointly explained by $X_{iZ}i$ and X_i (excluding the effect of Z_i). $R^2_{mo\Delta}$ ranges from [0, 1].

The second set of indicators is the V^2_R effect size indicators, including V^2_{R21}, V^2_{R22}, and V^2_{R23}:

$$V^2_{R21} = \frac{Var(\gamma_{11}X_{iZ}i)}{Var(\gamma_{10}X_i + \gamma_{11}X_{iZ}i + \varepsilon_{1i}X_i)}$$

$$V^2_{R22} = \frac{Var(\gamma_{11}X_{iZ}i)}{Var(\gamma_{10}X_i + \gamma_{11}X_{iZ}i)}$$

$$V_{R23}^2 = \frac{Var(\gamma_{11} X_{iZ} i)}{Var(\gamma_{10} X_i + \gamma_{11} X_{iZ} i) + Var(\varepsilon_{1i} X_i + e_i)}$$

The numerators of these three indicators are identical, representing the variance in Y_i explained by $X_{iZ}i$, but the denominators differ. The denominators of V_{R22}^2 and V_{R23}^2 are subsets of the denominator of V_{R21}^2, so both V_{R22}^2 and V_{R23}^2 are larger than V_{R21}^2. V_{R21}^2 represents the proportion of variance in Y_i explained by $X_{iZ}i$ relative to the variance in Y_i related to X_i. V_{R22}^2 represents the proportion of variance in Y_i explained by $X_{iZ}i$ relative to the variance in Y_i explained by X_i. V_{R23}^2 represents the proportion of variance in Y_i explained by $X_{iZ}i$ relative to the sum of variance in Y_i explained by $X_{iZ}i$ and variance in Y_i not explained by $X_{iZ}i$, similar to the effect size η_p^2 in ANOVA. V_{R21}^2 ranges from [0, 1], but V_{R22}^2 and V_{R23}^2 may exceed 1. In such cases, it is recommended to use alternative effect size indicators or center both the independent variable X and moderator Z before conducting moderation analysis and computing effect sizes.

3.2.3. Discussion of Effect Sizes

Effect sizes should have three properties: independence from measurement units, independence from sample size, and monotonicity. The multiple regression moderation effect size index ΔR^2 possesses all three properties (Wen et al., 2016). $R_{\beta 1}^2$ has the properties of being independent of measurement units and sample size, but it is a monotonic function of the conditional coefficient γ_{11} (Liu & Yuan, in press). When both independent variable X and moderator Z are centered, V_{R21}^2, V_{R22}^2, and V_{R23}^2 have the properties of being independent of measurement units and sample size, though Liu and Yuan did not discuss their monotonicity. Because $a^2x^2/(a^2x^2 + b^2)$ is a monotonic function of x^2 when a and b remain constant, V_{R21}^2, V_{R22}^2, and V_{R23}^2 all have monotonicity. Liu and Yuan noted that among the three V_R^2 effect size indicators, V_{R21}^2 best aligns with the definition of moderation. Therefore, we recommend centering both X and Z before computing V_R^2 effect sizes. When computing V_R^2, we suggest first calculating V_{R21}^2 and reporting the denominator variance $Var(\beta_{1i} X_i)$ to facilitate understanding of V_{R21}^2's practical significance. Then, depending on specific research needs, report V_{R22}^2 and V_{R23}^2.

4. Analysis Procedure

When faced with a moderation analysis task, how should researchers proceed? Based on the preceding discussion, we propose a comprehensive moderation analysis procedure (see Figure 3 [Figure 3: see original paper]):

Step 1. Conduct moderation analysis using a two-level regression model and test whether γ_{11} (see equation (3b)) is significant. If γ_{11} is significant, report the two-level regression model moderation analysis results and effect size. If not, proceed to Step 2.

Step 2. Test whether the variance σ_1^2 (see equation (3b)) is significant. If yes, then error variance is definitely heteroscedastic; report the two-level regression model moderation analysis results and effect size. If no, this means we cannot reject the null hypothesis that $\sigma_1^2 = 0$, and we cannot conclude homoscedasticity (i.e., $\sigma_1^2 = 0$). Therefore, proceed to Step 3.

Step 3. Conduct moderation analysis using multiple regression and test whether the BIC of the two-level regression model is smaller than the BIC of the multiple regression model. If yes, the two-level regression model fits the data better (Yuan et al., 2014); report the two-level regression model analysis results and effect size. If no, report the multiple regression analysis results and effect size ΔR^2.

Figure 3. Flowchart of Moderation Analysis Procedure

5. Example

We now demonstrate two-level regression model moderation analysis using an empirical example. This study examines the moderating role of callous-unemotional traits (Z) in the relationship between childhood maltreatment (X) and adolescents' cyberbullying behavior (Y). The variables and data (N = 940) are from Fang et al. (2020). All variables were standardized before analysis. We conducted two-level regression model moderation analysis using Mplus 8.2 (see Appendix for Mplus program). Note that first, by setting each Level 2 cluster to contain only one Level 1 subject (Mplus statement: CLUSTER = ID), Mplus treats single-level data as two-level data, enabling two-level regression model moderation analysis (Liu et al., 2020). Second, Mplus uses maximum likelihood estimation with robust standard errors (MLR) by default; Liu et al. (2020) recommend using Bayesian estimation (available in Mplus 8.3 and above). Third, Mplus assumes $\sigma_{01} = 0$ by default because when errors are uncorrelated ($\sigma_{01} = 0$), more accurate γ_{11} and σ_1^2 values can be obtained (Yuan et al., 2014; Liu et al., 2020, in press; Liu & Yuan, in press).

Results of the two-level regression model moderation analysis are shown in Table 1 . Since σ_1^2 is significant ($\hat{\sigma}_1^2$ = .29, Z = 3.46), we report the two-level regression model moderation analysis results. The estimated coefficient γ_{11} is significant ($\hat{\gamma}_{11}$ = .11, t = 2.89, p = .004), indicating that the moderation effect is significantly different from 0. The moderation effect size obtained from Mplus is $R^2_{\beta 1} = .04$, with a 95% confidence interval of [0, 0.085], indicating that 4% of the variance in β_{1i} can be explained by moderator Z. Effect sizes $R^2_{mo\Delta}$, V^2_{R21}, V^2_{R22}, and V^2_{R23} were computed using R software (Liu & Yuan, in press). Results show $R^2_{mo\Delta} = .64$, indicating that the variance in Y_i uniquely explained by $X_{iZ}i$ accounts for 64% of the variance jointly explained by $X_{iZ}i$ and X_i; $V^2_{R21} = .043$, indicating that the variance in Y_i explained by $X_{iZ}i$ accounts for 4.3% of the variance in Y_i related to X_i (which is .355); $V^2_{R22} = .229$, indicating that the variance explained by $X_{iZ}i$ accounts for 22.9% of the variance in Y_i explained by X_i (which is .067); and $V^2_{R23} = .05$, indicating that the variance explained by $X_{iZ}i$ accounts for 5% of the sum of variance in Y_i explained by $X_{iZ}i$ and variance not explained by $X_{iZ}i$ (which is .303).

Table 1. Results of Two-Level Regression Model Moderation Analysis from Mplus

Parameter	Estimate	SE	95% CI
γ_{00}	-0.69	0.41	[-1.49, 0.11]
γ_{01}	4.13***	0.77	[2.62, 5.64]
γ_{10}	5.36***	0.76	[3.87, 6.85]
γ_{11}	2.89**	1.00	[0.93, 4.85]
σ_1^2	3.46**	1.24	[1.03, 5.89]
σ_e^2	5.73***	0.26	[5.22, 6.24]

Note: ** $p < .01$; *** $p < .001$.

6.1. Robust Moderation Analysis

When data contain outliers or heavy tails, two-level regression model moderation analysis results can be biased (Yuan et al., 2014; Yang & Yuan, 2016). To address this issue, Yang and Yuan (2016) proposed two robust methods for two-level regression model moderation analysis. One is maximum likelihood estimation based on t-distribution weights, and the other is M-estimation based on Huber weights, where "M" represents maximum likelihood estimation.

The basic idea of robust methods is to assign appropriate weights to each data point, giving smaller weights to data far from the center (such as outliers), thereby reducing the influence of outliers on moderation effect estimates and obtaining more accurate estimates. Simulation studies by Yang and Yuan (2016) showed that when errors are non-normal (e.g., heavy-tailed), both robust methods outperform maximum likelihood estimation based on normal distribution in moderation analysis; when errors are normally distributed, both robust methods perform comparably to maximum likelihood estimation.

Applying robust two-level moderation analysis to the example, results show that the estimated coefficient γ_{11} is significant for both robust methods (Table 2), indicating that the moderation effect is significantly different from 0. However, both coefficient estimates are smaller than those from maximum likelihood estimation with robust standard errors ($\hat{\gamma}_{11} = .11$).

Table 2. Results of Robust Two-Level Regression Model Moderation Analysis

Method	γ_{11}	SE	95% CI
Huber weights	0.09	0.03	[0.03, 0.15]
t-distribution weights	0.08	0.03	[0.02, 0.14]

6.2. Moderation Analysis with Latent Variables

A limitation of observed variable moderation analysis is the assumption that all variables are measured without error. When measurement error is substantial, this can lead to underestimation of moderation effects. The primary advantage of conducting moderation analysis using structural equation modeling (SEM) is the ability to specify latent variables, effectively control for measurement error, and accurately estimate moderation effects.

Specifically, assuming latent variable ξ_x is measured by p_x indicators $x_1, x_2, ..., x_{p_x}$, latent variable ξ_z is measured by p_z indicators $z_1, z_2, ..., z_{p_z}$, and latent variable η_y is measured by p_y indicators $y_1, y_2, ..., y_{p_y}$, Liu et al. (2020) noted that two-level regression model SEM moderation analysis simply involves replacing observed variables Y, X, and Z in equations (2) and (3) with latent variables η_y, ξ_x, and ξ_z, respectively, and replacing the rectangular boxes (representing observed variables) with ellipses (representing latent variables) in Figure 2. The moderation effect size indicator $R^2_{\beta 1}$ only requires replacing $Var(Z_i)$ in equation (8) with $Var(\xi_z)$. Simulation studies by Liu et al. (2020) found that when error variances are heteroscedastic, two-level regression model SEM moderation analysis provides more accurate moderation effect estimates and better performance in terms of confidence interval coverage (closer to 95%) and Type I error rate (closer to .05) compared to single-level SEM moderation analysis (including product indicator methods and latent moderated structural equations). When error variances are homoscedastic, results from two-level regression model SEM moderation analysis are comparable to single-level analysis. Liu et al. provided Mplus programs for two-level regression model SEM moderation analysis. Notably, it remains unclear whether $R^2_{mo\Delta}$ and V^2_R indices are applicable to two-level regression model SEM moderation analysis (Liu & Yuan, in press).

6.3. Moderated Mediation Analysis Based on a Two-Level Regression Model

Currently, the integration of moderation and mediation has become a trend in social sciences. Liu et al. (in press) proposed a method for moderated mediation analysis based on a two-level regression model. For example, moderated mediation analysis where the latter mediation path (M→Y) is moderated can be expressed as:

Level 1:

$$Y_i = c'_0 + b_{iM} i + \beta_{00} X_i + e_{Yi}$$

$$M_i = a_0 + a_1 X_i + e_{Mi}$$

Level 2:

$$b_i = \gamma_{b0} + \gamma_{b1} Z_i + \varepsilon_{bi}$$

If γ_{b1} is significantly different from 0 (i.e., the confidence interval for γ_{b1} does not contain 0), it indicates that the mediation effect is moderated. Liu et al. (in press) defined the effect size of moderated mediation as the proportion of variance in the mediation effect $a_1 b_i$ that can be explained by Z_i, that is:

$$R^2_{med} = \frac{Var(a_1\gamma_{b1}Z_i)}{Var(a_1\gamma_{b1}Z_i + a_1\varepsilon_{bi})}$$

Similarly, readers can easily derive the equations for moderated mediation analysis where the former mediation path (X→M) is moderated:

Level 1:

$$M_i = a_{0i} + a_1 X_i + e_{Mi}$$
$$Y_i = c'_0 + b_1 M_i + \beta_{00} X_i + e_{Yi}$$

Level 2:

$$a_{0i} = \gamma_{a0} + \gamma_{a1} Z_i + \varepsilon_{ai}$$

If γ_{a1} is significantly different from 0 (i.e., the confidence interval for γ_{a1} does not contain 0), it indicates that the mediation effect is moderated. The effect size of moderated mediation is defined as the proportion of variance in the mediation effect $a_{0i} b_1$ that can be explained by Z_i, that is:

$$R^2_{med} = \frac{Var(b_1\gamma_{a1}Z_i)}{Var(b_1\gamma_{a1}Z_i + b_1\varepsilon_{ai})}$$

Liu et al. provided Mplus programs for two-level regression model moderated mediation analysis and effect size computation. Simulation studies by Liu et al. (in press) found that when error variances are heteroscedastic, two-level regression model moderated mediation analysis provides more accurate conditional mediation effect estimates and better performance in confidence interval coverage (closer to 95%) and Type I error rate (closer to .05) compared to single-level moderated mediation models. When error variances are homoscedastic, results from two-level regression model moderated mediation analysis are comparable to single-level analysis.

6.4. Mediated Moderation Analysis Based on a Two-Level Regression Model

In addition to moderated mediation analysis, Liu et al. (2021) also proposed a method for mediated moderation analysis based on a two-level regression model (where the moderating effect of Z on the X→Y relationship is transmitted through mediator M). Mediated moderation analysis based on a two-level regression model (Figure 4 [Figure 4: see original paper]) can be expressed as:

Level 1:

$$Y_i = c_{0i} + c_{1i}X_i + e_{Yi}$$
$$M_i = a_0 + a_1 Z_i + e_{Mi}$$

Level 2:

$$c_{0i} = \gamma_{c00} + \gamma_{c01} Z_i + \varepsilon_{0i}$$
$$c_{1i} = \gamma_{c10} + \gamma_{c11} Z_i + \gamma_{c12} M_i + \varepsilon_{1i}$$

Compared to two-level regression model moderation analysis (equations (2) and (3)), mediated moderation analysis only adds equation (13) and includes mediator M_i in equation (12). Substituting equations (12a) and (12b) into equation (11) yields:

$$Y_i = \gamma_{c00} + \gamma_{c01} Z_i + \gamma_{c10} X_i + \gamma_{c11} X_{iZ} i + \gamma_{c12} X_{iM} i + (\varepsilon_{0i} + \varepsilon_{1i} X_i + e_{Yi})$$

Because this is single-level data (equation (14) contains only subscript i), ε_{0i} and e_{Yi} cannot be distinguished. Substituting equation (13) into equation (14) and rearranging gives:

$$Y_i = (\gamma_{c00} + \gamma_{c12} a_0) + (\gamma_{c01} + \gamma_{c12} a_1) Z_i + \gamma_{c10} X_i + (\gamma_{c11} + \gamma_{c12} a_1) X_{iZ} i + (\varepsilon_{0i} + \varepsilon_{1i} X_i + \gamma_{c12} e_{Mi} X_i + e_{Yi})$$

$\gamma_{c11} + \gamma_{c12} a_1$ represents the moderating effect of Z on the X→Y relationship, where γ_{c11} is the direct moderating effect and $\gamma_{c12} a_1$ is the indirect moderating effect of Z on the X→Y relationship through M (i.e., mediated moderation). If γ_{c12} is significantly different from 0 (i.e., the confidence interval for γ_{c12} does not contain 0), it indicates that the moderating effect of Z on the X→Y relationship is mediated by M. Liu et al. (2021) defined the effect size of mediated moderation as the proportion of variance in the effect of independent variable X on dependent variable Y (c_{1i}) that can be explained by the mediated moderation effect $\gamma_{c12} a_1$, that is:

$$R^2_{medmod} = \frac{Var(\gamma_{c12} a_1 Z_i)}{Var(\gamma_{c11} Z_i + \gamma_{c12} a_1 Z_i + \varepsilon_{1i})}$$

Liu et al. (2021) provided Mplus programs for two-level regression model mediated moderation analysis and effect size computation. Their simulation studies found that when error variances are heteroscedastic (i.e., $\sigma_1^2 \neq 0$), two-level regression model mediated moderation analysis provides more accurate parameter and effect size estimates and Type I error rates closer to .05 compared to single-level mediated moderation models (which remove ε_{1i} from equation (12)). As σ_1^2 increases, the advantages of the two-level regression model become more pronounced. When error variances are homoscedastic (i.e., $\sigma_1^2 = 0$), results from two-level regression model mediated moderation analysis are comparable to single-level analysis.

7. Conclusion

In response to the limitations of multiple regression moderation analysis and its effect size ΔR^2 (Yuan et al., 2014; Liu & Yuan, in press), a better approach is to use two-level regression models for moderation analysis with corresponding new effect size indices. We have proposed a comprehensive moderation analysis procedure and demonstrated through an example how to conduct two-level regression model moderation analysis and compute its effect sizes. We also discussed future developments in two-level regression model moderation analysis. However, this paper has some limitations that require further discussion and extension. For example, this paper only involves moderation analysis with one independent variable, one moderator, and one dependent variable, but the methods introduced here are equally applicable to situations with multiple independent variables and multiple moderators. Yuan et al. (2014) used two-level regression models to analyze moderation effects with two independent variables and three moderators. Liu et al. (in press) used two-level regression models to analyze moderated mediation effects where two moderators separately moderated the front and back paths. Methodological advances provide researchers with opportunities to deepen their understanding and application of two-level regression models for moderation analysis, and we believe that continued research will enhance our understanding of moderation analysis using two-level regression models.

References

Fang, J., Lai, X., & Luo, W. (2020). The effect of childhood maltreatment on adolescents' cyberbullying: The moderating roles of callous-unemotional traits and gender. *Chinese Journal of Clinical Psychology*, 28(5), 991–994.

Fang, J., Wen, Z., Liang, D., & Li, N. (2015). Moderation analysis based on multiple regression. *Journal of Psychological Science*, 38(3), 715–720.

Liu, H., Yuan, K.-H., & Gan, K. (2021). Extension of mediated moderation models and measurement of their effect sizes. *Acta Psychologica Sinica*, 53(3), 322–338.

Wen, Z., Fan, X., Ye, B., & Chen, Y. (2016). On the rationality of mediation effect size from the perspective of properties that effect sizes should have. *Acta Psychologica Sinica*, 48(4), 435–443.

Wen, Z., Hou, J., & Zhang, L. (2005). Comparison and application of moderation and mediation effects. *Acta Psychologica Sinica*, 37(2), 268–274.

Wen, Z., & Liu, H. (2020). *Mediation and moderation effects: Methods and applications*. Beijing: Educational Science Publishing House.

Wen, Z., & Ye, B. (2014). Test methods for moderated mediation models: Competitors or supplements? *Acta Psychologica Sinica*, 46(5), 714–726.

Aguinis, H., Petersen, S. A., & Pierce, C. A. (1999). Appraisal of the homogeneity of error variance assumption and alternatives to multiple regression for estimating moderating effects of categorical variables. *Organizational Research Methods*, 2(4), 315–339.

Aguinis, H., Beaty, J. C., Boik, R. J., & Pierce, C. A. (2005). Effect size and power in assessing moderating effects of categorical variables using multiple regression: A 30-year review. *Journal of Applied Psychology*, 90(1), 94–107.

Hayes, A. F., & Montoya, A. K. (2017). A tutorial on testing, visualizing, and probing interaction involving a multicategorical variable in linear regression analysis. *Communication Methods and Measures*, 11, 1–30.

Liu, H., & Yuan, K-H. (in press). New measures of effect size in moderation analysis. *Psychological Methods*.

Liu, H., Yuan, K-H., & Liu, F. (2020). A two-level moderated latent variable model with single level data. *Multivariate Behavioral Research*, 55(6), 873–893.

Liu, H., Yuan, K.-H., Wen. Z. (in press). Two-level moderated mediation models with single level data and new measures of effect sizes. *Behavior Research Methods*.

Yang, M., & Yuan, K.-H. (2016). Robust methods for moderation analysis with a two-level regression model. *Multivariate Behavioral Research*, 51(6), 757–771.

Yuan, K.-H., Cheng, Y., & Maxwell, S. (2014). Moderation analysis using a two-level regression model. *Psychometrika*, 79(4), 701–732.

Appendix: Mplus Program for Two-Level Regression Model Moderation Analysis

```
DATA: FILE = 2.txt;
VARIABLE: NAMES ARE ID x z y;    !Variable names in 2.txt
          USEVARIABLES = x z y xz;
          CLUSTER = ID;    !ID is subject identifier
ANALYSIS: TYPE IS TWOLEVEL RANDOM;
          ALGORITHM = INTEGRATION;
MODEL: %WITHIN%
       XZ(r11);    !gamma11 value
       X with Z;
       X with Y;
       %BETWEEN%
       [b];                !gamma10 value
       b(sig2_ei1);  !sigma1 squared
MODEL CONSTRAINT:
       NEW(var_z R2);
       var_z=1;    !Variance of Z is 1 due to standardization
       R2=(r11*r11*var_z)/((r11*r11*var_z)+sig2_ei1);    !Compute effect size
```

```
OUTPUT: CINTERVAL;
```

Note: This program uses MLR estimation by default. For Bayesian estimation, change the ANALYSIS command to:

```
ANALYSIS: ESTIMATOR IS BAYES;
          POINT = MEAN;    !Use posterior mean as point estimate
```

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv — Machine translation. Verify with original.