STRUCTURAL EQUATION MODELING
2024, VOL. 31, NO. 2
ISSN: 1070-5511 | E-ISSN: 1532-8007

A Note on Evaluating the Moderated Mediation Effect

Chi Kit Jacky Ng[a], Lok Yin Joyce Kwan[b] and Wai Chan[c]

[a]The Hong Kong Polytechnic University; [b]The Education University of Hong Kong; [c]The Chinese University of Hong Kong

ABSTRACT
In the past decade, moderated mediation analysis has been extensively and increasingly employed in social and behavioral sciences. With its widespread use, it is particularly important to ensure the moderated mediation analysis will not bring spurious results. Spurious effects have been studied in both mediation and moderation analysis, but this issue remains unexplored in moderated mediation analysis. To fill this gap, we examined the conditions under which a spurious moderated mediation effect in a dual stage moderated mediation model might occur. Specifically, with a hypothetical example and three theorems, we illustrated how the index of moderated moderated mediation may conclude a moderated mediation effect which does not actually exist. As a remedy to rule out the spurious results, we proposed two methods which are simple and easy to implement. Based on the simulation results, we offer researchers some practical guidelines to apply the methods in empirical research.

KEYWORDS
Index of moderated moderated mediation; moderated mediation analysis; spurious effect

1. A Note on Evaluating the Moderated Mediation Effect

To capture the complexity of a social and psychological phenomenon, statistical models that involve multiple variables are often adopted. Mediation and moderation models are two of the most widely used statistical models in social and behavioral sciences (Aguinis et al., 2005, 2017; MacKinnon & Fairchild, 2009; Preacher, 2015). They help to identify the underlying mechanisms and the boundary conditions of a relationship (Aiken & West, 1991; Baron & Kenny, 1986; Cohen et al., 2003). Both of these models have been frequently tested over the past thirty years. There is also another statistical model that has been widely and increasingly tested in the past decade—a moderated mediation model.

Using the PsycINFO database, a literature search was conducted on various disciplines including social, health, organizational and behavioral sciences, covering the years between 2006 and 2020. In the past decade, 34 papers tested a moderated mediation model in 2010, 175 papers in 2015, and 551 papers in 2020, clearly revealing an ever-increasing trend for testing a moderated mediation model (Figure 1). Moreover, according to the Google Scholar Citation Index, the key methodological papers that discuss the moderated mediation model have been cited 22,971 times (e.g., Edwards & Lambert, 2007; Fairchild & MacKinnon, 2009; Hayes, 2015, 2018; Muller et al., 2005; Preacher et al., 2007). Work discussing a user-friendly macro in SPSS, SAS, and R to test a moderated mediation model (PROCESS; Hayes, 2022; Hayes & Rockwood, 2017, 2020; Hayes et al., 2017; Igartua & Hayes, 2021) has been cited 53,479 times. Recent research continues to initiate advancements in moderated mediation, including the extension to latent variables (Chen & Cheng, 2014; Cheung & Lau, 2017), Bayesian estimation (Wang & Preacher, 2015; Wedel & Dong, 2022), multilevel contexts (Bauer et al., 2006; Kim & Hong, 2020), differentiation from mediated moderation (Kwan & Chan, 2018; Ng et al., 2019), effect size (Liu et al., 2022), causal assumptions (Loeys et al., 2016), robust methods (Yuan & Gomer, 2021), component vs. index approaches (Yzerbyt et al., 2018) and power analysis (Thoemmes et al., 2010). These figures indicate a widespread and emerging interest in moderated mediation analysis across disciplines.

If a statistical model is extensively utilized, it is quite important to ensure that the model will not conclude a spurious effect. A spurious effect refers to a null effect that is incorrectly found to be significant due to statistical artifacts (Lubinski & Humphreys, 1990). Under some conditions, a statistical model may indicate the presence of an effect that does not exist, yielding a spurious effect. The issue of spurious effect has been studied in both mediation (e.g., Lemmer & Gollwitzer, 2017; Shrout & Bolger, 2002) and moderation models (e.g., Harring et al., 2015; Kromrey & Foster-Johnson, 1999). Nonetheless, this issue remains unexplored in a moderated mediation model. Therefore, we aim at investigating the spurious moderated mediation effect. In general, there are two basic moderated mediation models, namely the first stage (Figure 2A) and second stage (Figure 2B) moderated mediation models. Based on this, a more general model can be formulated by combining the first stage and the second stage moderated mediation effects, yielding a dual stage moderated mediation model

(Figure 2C). In this article, we aim at studying the spurious effect in this model.

In the following sections, we first discuss the formulation of a dual stage moderated mediation model. With a hypothetical example, we then illustrate the methodological problem in the model through three theorems and propose two methods (viz., the inverse- and log-transformation methods) to rule out the spurious effect. Using simulated data, we examine the statistical performance of the two methods and provide practical guidelines for applied researchers to use to check for the spurious moderated mediation effect based on some recommended cut-off values.

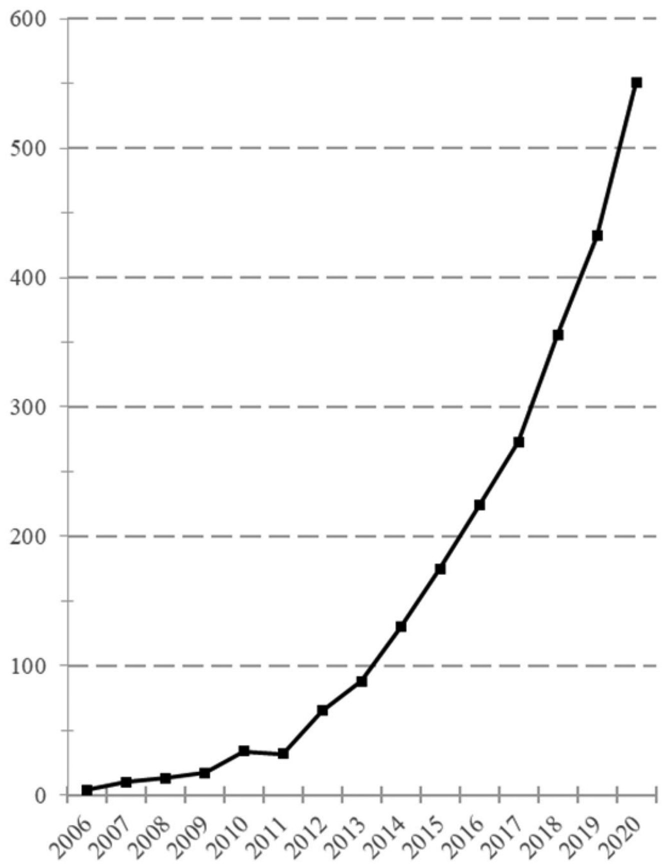

Figure 1. An exponentially increasing trend in testing a moderated mediation model in applied research from 2006 to 2020.

2. Moderated Mediation Models

James and Brett (1984) released the additivity requirement of mediation by including a moderator into the mediation model and attempted to demonstrate the existence of a moderated mediation model. After systematically formulating the tests of mediation and moderation models, Baron and Kenny (1986) also presented an example combining both mediation and moderation models, demonstrating the possibility of establishing a moderated mediation model. Built upon the previous work, Muller et al. (2005) comprehensively laid out the model equations underlying a moderated mediation process and introduced a fundamental equality to decompose a moderated overall effect into a moderated mediation effect and a moderated direct effect. In general, moderated mediation refers to a mediating process being moderated by one or more variables and can be manifested in different forms. Moderated mediation varies in different forms, depending mostly on the number of moderators and the paths being moderated. Edwards and Lambert (2007) and Preacher et al. (2007) discussed a number of common forms and highlighted the mathematical expression of the moderated mediation effect (also known as the conditional indirect effect) for each form (see also Little et al., 2007). In general, the two most common forms of moderated mediation are the first stage (Figure 2A) and second stage (Figure 2B) moderated mediation models in which either the first stage or second stage of a mediating process varies as a function of a moderator (see also Models B and C in Edwards & Lambert, 2007 and Little et al., 2007; Models 2 and 3 in Preacher et al., 2007). The first stage and second stage moderated mediation models can be combined as another common form—a dual stage moderated mediation model (Figure 2C), which is also the model of interest in this

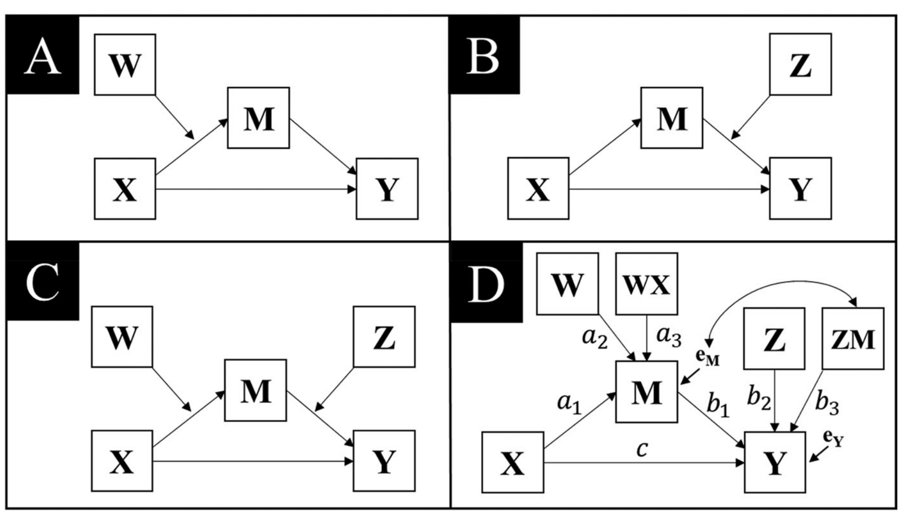

Figure 2. Conceptual models of the first stage (A), Second stage (B), and Dual stage (C) Moderated mediation, and the working model of the dual stage moderated mediation (D).

article (see also Model D in Little et al., 2007; Model 4 in Preacher et al., 2007).

2.1. Dual Stage Moderated Mediation Model

In a dual stage moderated mediation model, the first and second stages of a mediating process are influenced separately by two moderators. This model is represented with two model equations (see Figure 2D):

$$M = a_0 + a_1X + a_2W + a_3WX + e_M, \quad (1)$$

$$Y = b_0 + b_1M + b_2Z + b_3ZM + cX + e_Y. \quad (2)$$

Combining Equations (1) and (2), we have

$$\begin{aligned} Y &= b_0 + b_1(a_0 + a_1X + a_2W + a_3WX + e_M) + b_2Z \\ &\quad + b_3Z(a_0 + a_1X + a_2W + a_3WX + e_M) + cX + e_Y \\ &= b_0 + a_0b_1 + a_1b_1X + a_2b_1W + a_3b_1WX + b_1e_M \\ &\quad + b_2Z + a_0b_3Z + a_1b_3ZX + a_2b_3WZ + a_3b_3WZX \\ &\quad + b_3Ze_M + cX + e_Y \\ &= [(b_0 + a_0b_1) + a_2b_1W + (b_2 + a_0b_3)Z + a_2b_3WZ] \\ &\quad + [(a_1b_1 + c) + (a_3b_1W + a_1b_3Z + a_3b_3WZ)]X \\ &\quad + (b_1e_M + b_3Ze_M + e_Y) \end{aligned} \quad (3)$$

Based on the reduced form Equation (3), one can express the unconditional, conditional, and total effect of X on Y as follows (see also Hayes, 2018; Preacher et al., 2007):

(i) the unconditional (indirect and direct) effect of X on Y is $a_1b_1 + c$,
(ii) the conditional (indirect) effect of X on Y (conditional on W and Z) is

$$\gamma(W, Z) = a_3b_1W + a_1b_3Z + a_3b_3WZ, \quad (4)$$

(iii) and the total (unconditional and conditional) effect of X on Y is

$$T_{Y|X} = (a_1b_1 + c) + \gamma(W, Z) = c + (a_1 + a_3W)(b_1 + b_3Z). \quad (5)$$

2.2. Index of Moderated Moderated Mediation

From Equation (4), the conditional indirect effect from X to Y through M is a multiplicative function of W and Z. Hayes (2018) proposed the product coefficient a_3b_3 as an index of moderated moderated mediation, which quantifies the rate of change in the moderation effect of W on the indirect effect as Z changes, or equivalently the rate of change in the moderation effect of Z on the indirect effect as W changes. In other words, if a_3b_3 is different from zero, one can conclude a moderated mediation effect (or a conditional indirect effect) by W and Z and it is expected that the mediation effect of X on Y through M varies across different combinations of W and Z. More importantly, as will be shown in Theorem 1 below, it is only the interaction effect of W and Z (a_3b_3) in Equation (4) that is scale independent with respect to the transformation in the two moderators W and Z. On the other hand, according to the principle of marginality (Nelder, 1977), the main effects of W and Z (viz., a_3b_1 and a_1b_3) are scale dependent and marginal to their interaction effect and therefore, they may not be interpreted in a meaningful way. Consequently, the index of moderated moderated mediation is believed to precisely describe the relationship between the size of the indirect effect and the two moderators W and Z.

2.3. An Illustrative Example

In social and behavioral sciences literature, a dual stage moderated mediation model is typically formulated by combining two moderation models. For instance, Li et al. (2015) first hypothesized a moderation effect in which the influence of caregiving demands on family-to-work conflict might depend on the perceived strain in the family. They also proposed that the support received from a supervisor might alter the effect of family-to-work conflict on one's mental health. Taken together, a dual stage moderated mediation model was formulated, testing if the psychological conflict mechanism between caregiving demands and mental health would be contingent on the psychological resources received from the family and supervisor. Similar examples in social and behavioral sciences can be found in Armstrong et al. (2014), Bamberger and Belogolovsky (2017), Dufour et al. (2020), Gopalan et al. (2022), Laran et al. (2011), Mitchell et al. (2015), Racine and Martin (2017), and Yaffe and Kark (2011).

Following the above scenario, we present a hypothetical example to illustrate the dual stage moderated mediation model and the index of moderated moderated mediation. We also use this example to illustrate the theorems discussed below. Given a population model of a dual stage moderated mediation (see Equations 1 and 2, and Figure 2D) where $a_0 = b_0 = 0$, $a_1 = a_2 = b_1 = b_2 = c = 0.4$, and $a_3 = b_3 = 0.2$, the effects of interest are the two moderation effects from $W \left(\begin{array}{c} W \\ \downarrow \\ X \to M \end{array} \right)$ and $Z \left(\begin{array}{c} Z \\ \downarrow \\ M \to Y \end{array} \right)$. As a_3 (b_3) is non-zero, W (Z) moderates the effect of X (M) on M (Y). A sample of 10 observations was generated to visualize the moderation patterns graphically (Table 1). As shown in Figure 3A, W exacerbates the positive effect of X on M such that the effects of X are 0.06, 0.41, and 0.78 when W is relatively small (−1.71), moderate (0.04), and large (1.90), respectively. As shown in Figure 3B, Z exacerbates the positive effect of M on Y (after accounting for the effect of X on Y) such that the effects of M are 0.03, 0.48, and 1.16 when Z is relatively small (−1.85), moderate (0.41), and

Table 1. Sample of observations in the hypothetical example.

	Raw variables					Transformed variables	
Obs	X	M	Y	W	Z	W'	Z'
1	0.52	−0.74	0.87	−1.88	3.80	0.12	5.80
2	0.95	−0.63	0.24	−1.71	0.41	0.29	2.41
3	1.28	−0.16	−0.03	−1.02	−1.29	0.98	0.71
4	0.14	−0.37	−0.52	−1.00	−1.30	1.00	0.70
5	2.06	0.06	0.30	−0.95	−1.34	1.05	0.66
6	1.69	0.71	0.06	0.04	−1.66	2.04	0.34
7	0.33	0.65	−0.55	1.11	−1.78	3.11	0.22
8	1.33	1.61	−0.13	1.61	−1.81	3.61	0.19
9	3.33	3.36	0.73	1.90	−1.82	3.90	0.18
10	2.33	3.27	0.29	2.70	−1.85	4.70	0.15

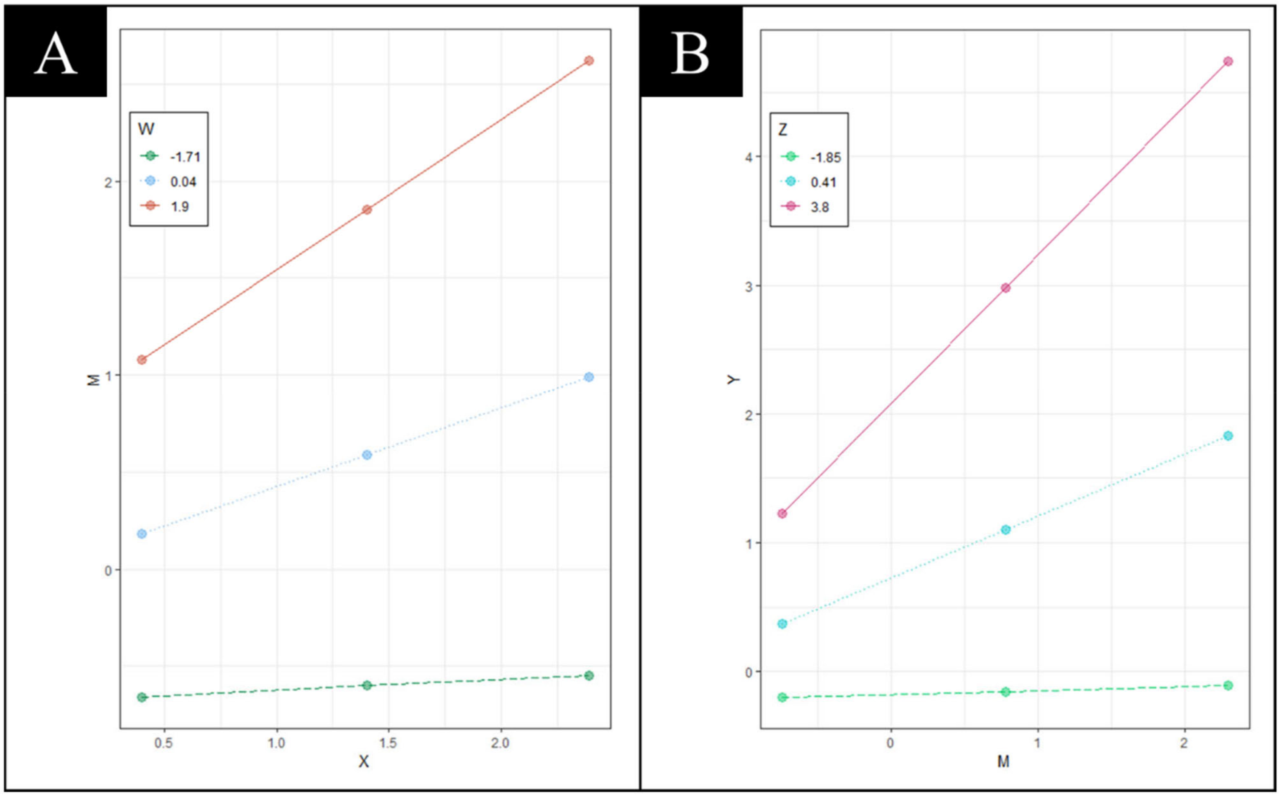

Figure 3. Simple slope plots for the first stage (A) and Second stage (B) of a Mediation process.

large (3.80), respectively. Taken together in a dual stage moderated mediation model, the effect of X on Y through M is conditional on different values of W and Z since the index of moderated moderated mediation ($a_3b_3 = 0.04$) is non-zero. Therefore, in this example, one can conclude the presence of a moderated mediation effect by W and Z. Critically, as the question of interest in this article, is it possible for the moderated mediation effect concluded to be spurious?

3. A Spurious Moderated Mediation Effect

The problem of spurious effect is a long-standing one in methodological research (Maxwell & Delaney, 1993; Simon, 1954). The problem of spurious effect has been investigated in both mediation and moderation analyses but this issue is rarely explored in moderated mediation analysis. Specifically, we will reveal that, under a certain condition, the index of moderated moderated mediation may indicate the presence of a moderated mediation effect that does not actually exist (i.e., a spurious moderated mediation effect) in a dual stage moderated mediation model.

In the following sections, three theorems will be used to illustrate the problem of the spurious moderated mediation effect. Theorem 1 demonstrates the effect of location shift of W and Z on the unconditional and conditional effect of X on Y. Theorem 2 demonstrates how the conditional effect of X on Y changes as a result of a location shift of W and Z. Theorem 3 further demonstrates the condition under which a spurious moderated mediation effect might occur. Furthermore, we assume $a_3 \neq 0$ and $b_3 \neq 0$ in all these theorems, so that W and Z moderate the first and second stages of the mediation pathway, respectively. The hypothetical example discussed above will then be used to exemplify the theorems. As a remedy, two methods are proposed to rule out the possibility of concluding a spurious moderated mediation effect.

Theorem 1. *A location shift of* $\boldsymbol{W}$ *and* $\boldsymbol{Z}$ *will change the unconditional effect of* $\boldsymbol{X}$ *on* $\boldsymbol{Y}$ *and the main effects of* $\boldsymbol{W}$ *and* $\boldsymbol{Z}$*, but it will not change the interaction effect of* $\boldsymbol{W}$ *and* $\boldsymbol{Z}$*.*

Proof. Consider a location shift transformation of W and Z such that $W^* = W + k$ and $Z^* = Z + m$, where k and m are two location shift constants. From Equation (5),

$$\begin{aligned} T_{Y|X} &= (a_1b_1 + c) + (a_3b_1W + a_1b_3Z + a_3b_3WZ) \\ &= a_1b_1 + c + a_3b_1(W^* - k) + a_1b_3(Z^* - m) \\ &\quad + a_3b_3(W^* - k)(Z^* - m) \\ &= a_1b_1 + c + a_3b_1W^* - a_3b_1k + a_1b_3Z^* - a_1b_3m \\ &\quad + a_3b_3W^*Z^* - a_3b_3kZ^* - a_3b_3mW^* + a_3b_3km \\ &= [(a_1 - a_3k)(b_1 - b_3m) + c] + [a_3(b_1 - b_3m)W^* \\ &\quad + (a_1 - a_3k)b_3Z^* + a_3b_3W^*Z^*] \\ &= [a_1^*b_1^* + c] + [a_3b_1^*W^* + a_1^*b_3Z^* + a_3b_3W^*Z^*], \end{aligned} \tag{6}$$

where $a_1^* = a_1 - a_3k$ and $b_1^* = b_1 - b_3m$.

From Equation (6), we can see that after the location shift transformation of W and Z, the unconditional (indirect and direct) effect of X on Y is

$$a_1^*b_1^* + c = (a_1 - a_3k)(b_1 - b_3m) + c, \tag{7}$$

and the conditional (indirect) effect of X on Y (conditional on W^* and Z^*) is

$$\begin{aligned}\gamma(W^*,Z^*) &= a_3b_1^*W^* + a_1^*b_3Z^* + a_3b_3W^*Z^* \\ &= a_3(b_1 - b_3m)W^* + (a_1 - a_3k)b_3Z^* + a_3b_3W^*Z^*.\end{aligned} \tag{8}$$

In other words, the unconditional indirect and direct effects are scale dependent with respect to the transformation of W and Z. For the conditional indirect effect, the main effect of W^* ($a_3b_1^*$) and Z^* ($a_1^*b_3$) are also scale dependent, while the interaction effect of W^* and Z^* (a_3b_3) is scale invariant with respect to the transformation.

Corollary (1). Define

$$k' = \frac{a_1}{a_3} \text{ and } m' = \frac{b_1}{b_3}, \tag{9}$$

and consider a location shift transformation of W and Z such that $W' = W + k'$ and $Z' = Z + m'$. According to Equations (7) to (9),

(i) the unconditional (indirect and direct) effect of X on Y is

$$\begin{aligned}a_1'b_1' + c &= (a_1 - a_3k')(b_1 - b_3m') + c \\ &= \left(a_1 - a_3\left(\frac{a_1}{a_3}\right)\right)\left(b_1 - b_3\left(\frac{b_1}{b_3}\right)\right) + c \\ &= c,\end{aligned}$$

which is indeed the direct effect of X on Y.

(ii) The conditional (indirect) effect of X on Y (conditional on W' and Z') is

$$\begin{aligned}\gamma(W',Z') &= a_3(b_1 - b_3m')W' + (a_1 - a_3k')b_3Z' + a_3b_3W'Z' \\ &= a_3b_3W'Z'.\end{aligned}$$

With an appropriate transformation, therefore, it is possible to eliminate the lower-order terms in $\gamma(W,Z)$ and the size of the conditional effect becomes a_3b_3. Given the fact that moderators in social sciences research are usually arbitrarily scaled (e.g., interval scale data), testing the hypothesis $a_3b_3 = 0$ appears to be a reasonable way to examine if the conditional indirect effect of X on Y exists.

Theorem 2. The conditional effect of $\boldsymbol{X}$ on $\boldsymbol{Y}$ will change by a constant after a location shift of $\boldsymbol{W}$ and $\boldsymbol{Z}$.

Proof. Consider a location shift transformation of W and Z such that $W^* = W + k$ and $Z^* = Z + m$. Based on Equations (4) and (8), the difference between $\gamma(W,Z)$ and $\gamma(W^*,Z^*)$ is

$$\begin{aligned}\Delta &= \gamma(W,Z) - \gamma(W^*,Z^*) \\ &= (a_3b_1W + a_1b_3Z + a_3b_3WZ) - (a_3b_1^*W^* + a_1^*b_3Z^* \\ &\quad + a_3b_3W^*Z^*) \\ &= a_3b_1W + a_1b_3Z + a_3b_3WZ - a_3(b_1 - b_3m)(W + k) \\ &\quad - (a_1 - a_3k)b_3(Z + m) - a_3b_3(W + k)(Z + m) \\ &= a_3b_1W + a_1b_3Z + a_3b_3WZ - a_3b_1W + a_3b_3mW \\ &\quad - a_3b_1k + a_3b_3km - a_1b_3Z + a_3b_3kZ - a_1b_3m \\ &\quad + a_3b_3km - a_3b_3WZ - a_3b_3kZ - a_3b_3mW - a_3b_3km \\ &= -a_3b_1k - a_1b_3m + a_3b_3km.\end{aligned} \tag{10}$$

From Equation (10), we can see that Δ is a constant as it does not depend on either W or Z.

Corollary (2). If $\gamma(W,Z) = \gamma$ where γ is a constant, then $\gamma(W^*,Z^*) = \gamma - \Delta$ will remain to be a constant.

For example, if $b_1 \neq 0$ and by choosing $k = \frac{-\gamma}{a_3b_1}$ and $m = 0$, then $\Delta = \gamma$ according to Equation (10). In other words, $\gamma(W^*,Z^*) = 0$. Similarly, if $a_1 \neq 0$ and by choosing $k = 0$ and $m = \frac{-\gamma}{a_1b_3}$, then $\Delta = \gamma$ and $\gamma(W^*,Z^*) = 0$.

Corollary (3). Suppose $\gamma(W,Z) = \gamma$ is a constant. By choosing $k = k'$ and $m = m'$,

$$\Delta = -a_3b_1\left(\frac{a_1}{a_3}\right) - a_1b_3\left(\frac{b_1}{b_3}\right) + a_3b_3\left(\frac{a_1}{a_3}\frac{b_1}{b_3}\right) = -a_1b_1. \tag{11}$$

Together with Corollary (1),

$$\begin{aligned}\gamma(W',Z') &= a_3b_3W'Z' = \gamma + a_1b_1 \\ W'Z' &= \frac{\gamma + a_1b_1}{a_3b_3} = \gamma' + k'm'\end{aligned} \tag{12}$$

where $\gamma' = \gamma/a_3b_3$. When the conditional effect is a constant, the product of W' and Z' will also be a constant,[1] or vice versa. Therefore, it is theoretically possible for the conditional effect of X on Y not to vary across different values of W and Z even when a_3b_3 is non-zero, yielding a spurious moderated mediation effect.

Theorem 3. When a specific nonlinear relationship exists between W and Z (that is, $Z = -m' + \frac{\delta}{W+k'}$ where δ is a non-zero constant), the conditional effect of X on Y will be equal to a constant, γ. That is, $\gamma(W,Z) = \gamma$.

Proof. Let

$$Z = -m' + \frac{\delta}{W + k'}, \tag{13}$$

where δ is a non-zero constant. From Equations (4) and (13), we have

$$\begin{aligned}\gamma(W,Z) &= a_3b_1W + a_1b_3Z + a_3b_3WZ \\ &= a_3b_1W + a_1b_3\left(-m' + \frac{\delta}{W + k'}\right) + a_3b_3W\left(-m' + \frac{\delta}{W + k'}\right) \\ &= a_3b_1W + \left(-a_1b_3m' + \frac{a_1b_3\delta}{W + k'}\right) + \left(-a_3b_3m'W + \frac{a_3b_3\delta W}{W + k'}\right) \\ &= a_3b_1W + \left(-a_1b_1 + \frac{a_1b_3\delta}{W + k'}\right) + \left(-a_3b_1W + \frac{a_3b_3\delta W}{W + k'}\right) \\ &= -a_1b_1 + \frac{a_3b_3\delta(W + k')}{W + k'} \\ &= -a_1b_1 + a_3b_3\delta = \gamma,\end{aligned} \tag{14}$$

where $\delta = \frac{\gamma + a_1b_1}{a_3b_3} = \gamma' + k'm'$, and $\gamma \neq -a_1b_1$ to ensure $\delta \neq 0$. When Equation (13) holds, the conditional indirect effect will no longer be "conditional," and the index of moderated moderated mediation can be misleading as it may conclude a spurious moderated mediation effect. Besides, when Equation (13) holds, W' and Z' are inversely proportional.

[1]In relation to this, $\gamma \neq -a_1b_1$, or $\gamma' \neq -k'm'$. Otherwise, $W'Z' = 0$, indicating that at least one of the moderators is a constant (i.e., $W = -k'$ or $Z = -m'$) which is rare in empirical research.

2.4. A Brief Discussion of Theorems

As a whole, three theorems were used to illustrate the problem of the spurious moderated mediation effect. In terms of identifying a spurious moderated mediation effect, Theorem 3 has the most direct implication since it highlights the condition where a spurious moderated mediation effect occurs (see Equation 13). Yet, we would like to emphasize that Theorems 1 and 2 are equally essential in illustrating the problem of the spurious moderated mediation effect.

On the one hand, Theorem 1 offers a mathematical justification of looking into a_3b_3 for the evidence of concluding a moderated mediation effect. As shown in Equation (4), the conditional effect in a dual stage moderated mediation model involves multiple components, and it is unclear which component should be focused to conclude a moderated mediation effect. Thus, through Theorem 1, we are able to locate the most appropriate conditional component (i.e., the scale-independent a_3b_3) and lay the foundation of studying which component may have a possibility in concluding a spurious moderated mediation effect.

On the other hand, Theorem 2 states that the moderated mediation effect will change by a constant given a location shift of the moderators. Subsequently, if a spurious moderated mediation effect is found under the original metrics of the moderators, it will remain spurious regardless of any location shift of the moderators. In social sciences research, variables are often measured in an interval scale where the choice of location is arbitrary and therefore, location shift of variables (e.g., linear transformation of interval variables) is indeed common in empirical research. In essence, Theorem 2 shows that the problem of spurious moderated mediation effect will persist even if the moderators undergo a linear transformation.

2.5. The Illustrative Example (Continued)

With the above hypothetical example, we will illustrate further the implications of the theorems. To recap the example, both a_3 and b_3 are non-zero, revealing the presence of two moderation effects in a dual stage moderated mediation model. More importantly, the product coefficient a_3b_3 is non-zero, indicating the presence of a moderated mediation effect where the mediation effect should vary depending on different values of W and Z. In the previous section, we ended the example by asking if it is possible for the concluded moderated mediation effect to be spurious. The answer is yes, the concluded moderated mediation effect is indeed spurious.

Based on Theorem 1, after transforming W and Z to W' and Z' where $k' = m' = \frac{0.4}{0.2} = 2$ (see the transformed variables in Table 1), the conditional indirect effect of X on Y by W' and Z' ($\gamma(W', Z')$) is $a_3b_3W'Z'$ (i.e., $0.2^2 \times W'Z'$). Based on Theorem 2, even though a_3b_3 is non-zero, it is possible that the conditional indirect effect of X on Y is not "conditional," showing a constant over different values of W' and Z'. For instance, the indirect effects of X on Y through M across different values of W' and Z' for the first three observations are

$$\gamma(W_1', Z_1') = 0.2^2 \times (0.12) \times (5.80) = 0.028;$$
$$\gamma(W_2', Z_2') = 0.2^2 \times (0.29) \times (2.41) = 0.028;$$
$$\gamma(W_3', Z_3') = 0.2^2 \times (0.98) \times (0.71) = 0.028.$$

The same pattern can also be seen in the conditional indirect effect of X on Y by W and Z in Equation (4):

$$\gamma(W_1, Z_1) = 0.4 \times 0.2 \times (-1.88) + 0.4 \times 0.2 \times (3.80) + 0.2 \times 0.2 \times (-1.88) \times (3.80) = -0.132$$

$$\gamma(W_2, Z_2) = 0.4 \times 0.2 \times (-1.71) + 0.4 \times 0.2 \times (0.41) + 0.2 \times 0.2 \times (-1.71) \times (0.41) = -0.132$$

$$\gamma(W_3, Z_3) = 0.4 \times 0.2 \times (-1.02) + 0.4 \times 0.2 \times (-1.29) + 0.2 \times 0.2 \times (-1.02) \times (-1.29) = -0.132.$$

A 3D plot was constructed to visualize the effect of X on Y (y-axis) by W (z-axis) and Z (x-axis). As shown in Figure 4, the indirect effect of X on Y shows a flat plane over W and Z, again indicating that the conditional indirect effect of X on Y is a constant. As a result, even though the index of moderated moderated mediation is non-zero, there is no moderated mediation effect in the example because the effect of X on Y is not conditional on W and Z (or W' and Z'). In this sense, a spurious moderated mediation effect is concluded.

Based on Theorem 3, if Equation (13) holds, W' and Z' are inversely proportional and a spurious moderated mediation effect results. This is evident in the example. Given $Z = -m' + \frac{\delta}{W+k'}$ where $\delta = \frac{\gamma + a_1b_1}{a_3b_3} = \frac{-.132+0.16}{0.04} = 0.70$, Equation (13) holds:

$$Z_1 = -2 + \frac{0.70}{W_1 + 2} = 3.80 \quad ;$$
$$Z_2 = -2 + \frac{0.70}{W_2 + 2} = 0.41 \quad ;$$
$$Z_3 = -2 + \frac{0.70}{W_3 + 2} = -1.29.$$

As demonstrated, a specific nonlinear relationship exists between W and Z and it contributes to the spurious moderated mediation effect in the example.

3. Two Proposed Methods

The hypothetical data set in Table 1 showed clearly that it is extremely difficult for one to tell if the moderated mediation effect is indeed spurious by simply "looking at the data". Similarly, the correlation between W and Z is $-.64$, a rather typical value that may not be informative enough to provide any insight for spotting the problem either (see also our discussion section below). As a result, we propose two empirical methods for evaluating if the conditional indirect effect is a constant (i.e., $\gamma(W, Z) = \gamma$), namely the *inverse-*

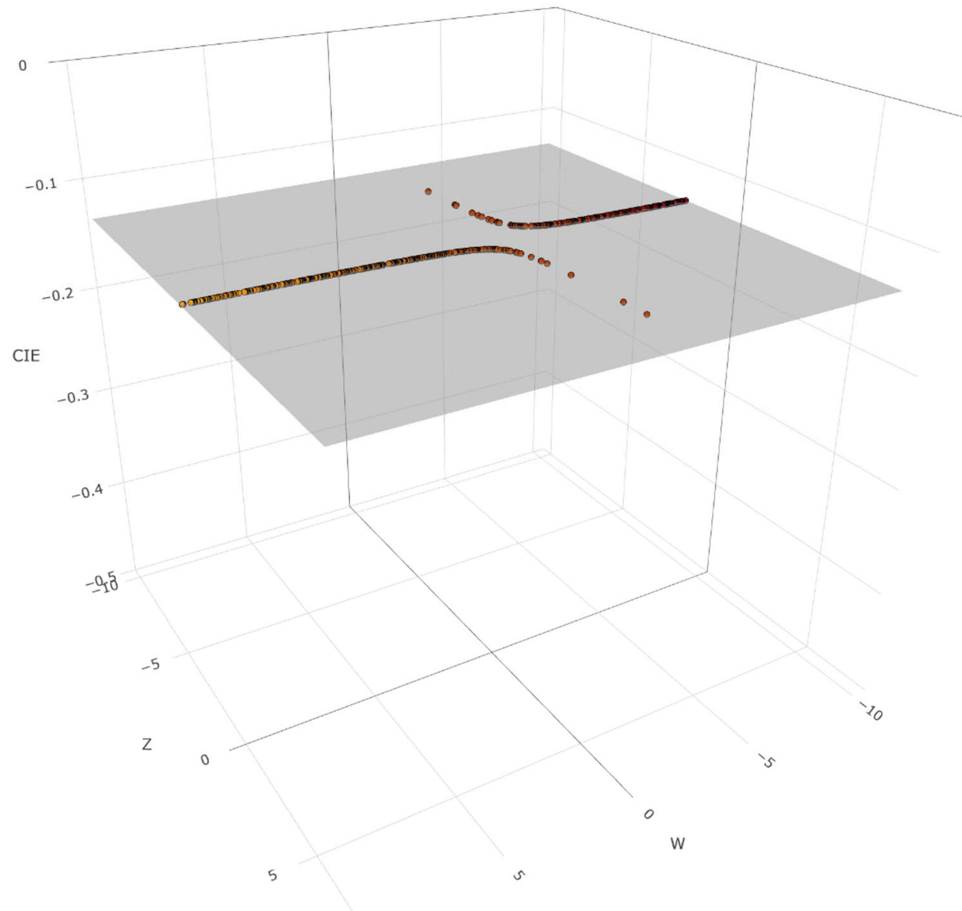

Figure 4. A 3D plot visualizing the conditional indirect effect of X on Y(y-axis) by W(z-axis) and Z (x-axis).

transformation method and the *log-transformation method*. For both methods, we first estimate k' and m' in Equation (9) by fitting a dual stage moderated mediation model to the sample data. Then, we transform W and Z to W' and Z' accordingly.

3.1. Inverse-Transformation Method

According to Equation (13), Z' and W' are inversely proportional across observations. The correlation between Z' and $\frac{1}{W'}$ (or W' and $\frac{1}{Z'}$) is expected to be +1.0 when δ is positive and −1.0 when δ is negative. As the population value of δ ($\delta = \frac{\gamma + a_1 b_1}{a_3 b_3}$) is generally unknown, we take the square for both sides in Equation (13):

$$Z'^2 = \frac{\delta^2}{W'^2}$$

As such, the correlation between Z'^2 and $\frac{1}{W'^2}$ (or W'^2 and $\frac{1}{Z'^2}$) is +1.0 regardless of the population value of δ. To illustrate the use of this method, consider our hypothetical example where a spurious moderated mediation effect is present. Two sets of correlations between one transformed moderator and the inverse of another transformed moderator were computed on the observations in Table 1. The correlations between Z'^2 and $\frac{1}{W'^2}$ as well as W'^2 and $\frac{1}{Z'^2}$ equal +1.0, indicating the presence of a spurious moderated mediation effect.

3.2. Log-Transformation Method

According to Equation (12), the product of W' and Z' will be a constant when the conditional indirect effect is a constant. In this method, we consider a log-transformation for the variables in Equation (12).[2]

To avoid negative values before transformation, we take the square for both sides:

$$\begin{aligned} W'^2 Z'^2 &= (\gamma' + k'm')^2 \\ \log(W'^2) + \log(Z'^2) &= \log\left((\gamma' + k'm')^2\right) = \log(\delta^2) \end{aligned} \tag{15}$$

From Equation (15), the sum of the log-transformed variables of W' and Z' is equal to a constant across observations. We call this the *log-ipsative property* of W' and Z'. Based on this property, the correlation between the log-transformed W'^2 and Z'^2 across the observations is expected to be −1.0 when the conditional indirect effect is a constant. Unlike the inverse-transformation method, only one set of correlations is required as the log-transformation is done on

[2]Although the common logarithm is used here, different types of logarithms (e.g., natural logarithm) are applicable to this method and will yield the same results.

both moderators. To illustrate the use of this method, we computed the correlation between the log-transformed W'^2 and Z'^2 in our example, which is equal to -1.0, again indicating the presence of a spurious moderated mediation effect.

Overall, if the conditional indirect effect is a constant at the population level, we expect to observe the correlations of $+1.0$ and -1.0 in the inverse- and log-transformation methods, respectively. Nonetheless, the estimated coefficients in a dual stage moderated mediation model (e.g., a_1, a_3, b_1, and b_3) are subject to sampling bias with a finite sample. As such, the product of the transformed moderators in Equation (12) may not be a constant in empirical studies even when the conditional indirect effect is a constant at the population level. Therefore, the correlations tested in both methods will generally deviate from $+1.0$ or -1.0 at the sample level (unlike the hypothetical example where the model coefficients are assumed to be known).

4. A Simulation Study Examining Inverse- and Log-Transformation Methods

Although the proposed methods are theoretically feasible, their empirical performance in revealing a spurious moderated mediation effect with sample data is unclear. Since both methods count on the estimate of the regression coefficients in the model (viz., $\widehat{a_1}$, $\widehat{a_3}$, $\widehat{b_1}$, and $\widehat{b_3}$), their performance may depend largely on the sample size in an empirical study. A Monte Carlo simulation study was conducted to evaluate the empirical performance of the two methods over a wide range of sample sizes. We also used the simulation results to establish some cut-offs for the correlations in both methods at the sample level.

4.1. Model Specification

A dual stage moderated mediation model was considered as the population model in this simulation study (Figure 2C). In this model, variables X, M, Y, W, and Z are regarded as the independent, mediating, dependent, first stage moderating, and second stage moderating variables respectively. This model is represented with two model equations (Figure 2D):

$$M = a_0 + a_1X + a_2W + a_3WX + e_M, \quad (16)$$

$$Y = b_0 + b_1M + b_2Z + b_3ZM + cX + e_Y. \quad (17)$$

The values of the population parameters in the two model equations were specified as follows: $a_0 = b_0 = 0$, $a_1 = a_2 = b_1 = b_2 = c = 0.40$, $a_3 = b_3 = 0.20$. These selected values are consistent with previous simulation studies (e.g., Cheung & Lau, 2017; Klein & Moosbrugger, 2000; Liu et al., 2022; Preacher et al., 2007; Wang & Preacher, 2015).

4.2. Study Design and Procedure

In this simulation study, we manipulated a wide range of sample sizes to draw the recommendations on the cut-off values for the correlations in both methods. Ten levels of sample size ranging from $N = 100$ to $N = 1000$ in an interval of 100 were manipulated to cover a reasonable range of sample sizes in social and behavioral sciences.

Given the model specified in (16) to (17), data was generated for each sample size condition in three steps. First, X was drawn from a normally distributed population with zero mean and unit variance, while e_M and e_Y were generated from a normally distributed population with zero mean(s) and their variances(s) were manipulated to yield 30% and 40% explained variance in M and Y respectively. The average sizes of the main effect (viz., a_1, a_2, b_1, b_2, and c) and the moderation effect (viz., a_3 and b_3) in Equations (16) to (17) were 18.0% (from 7.9% to 35.3%) and 8.4% (from 7.2% to 9.5%), respectively. These values fall within the common range reported in previous empirical studies (e.g., Champoux & Peters, 1987; Chaplin, 1991; Jaccard & Wan, 1995) and simulation studies (e.g., Foldnes & Hagtvet, 2014; Marsh et al., 2004; Moulder & Algina, 2002; Ng & Chan, 2020). Second, W was drawn from an F distribution with $df_1 = df_2 = 10$. In this case, the mean and variance of $\frac{1}{W}$ is well-defined in an F distribution. W was then converted to $\frac{1}{W}$, and Z was derived from $\frac{1}{W}$ through Equation (13) with $k' = m' = \frac{0.4}{0.2} = 2$ and $\delta = 2$. As the mean and the variance of $\frac{1}{W}$ is well-defined in an F distribution, the mean and variance of Z is also well-defined. Covariances among error terms as well as covariances between error terms and exogenous variables were fixed at zero. Finally, WX was computed by multiplying X by W, while M was derived by X, W, and WX through Equation (16). ZM was computed by multiplying M by Z, while Y was derived by X, M, Z, and ZM through Equation (17).

In addition to examining the performance of the proposed methods in conditions where a spurious moderated mediation effect occurs (as mentioned in the above data generation procedures), we included conditions where a genuine moderated mediation effect exists for comparison purposes. The simulation settings for these comparison conditions are the same as those in the spurious effect conditions, except for one setting in the second step of the data generation procedures mentioned above. Specifically, in the comparison conditions, Z was not derived by $\frac{1}{W}$ through Equation (13), but instead through a linear model $Z = -W + e$, where e is a normally distributed random variable with variance properly manipulated to give a correlation of -0.60 between Z and W. This value of correlation is the same as that in the spurious effect conditions. Overall, our simulation study examined ten sample size levels ranging from $N = 100$ to $N = 1000$ in both conditions with either a spurious or genuine moderated mediation effect.

R (*R* Development Core Team, 2021) was used for data generation and analysis. In each simulation condition, 1,000 replications were generated. In each replication, a dual stage moderated mediation model was tested using the *R* package of *lavaan* (Rosseel, 2012). Four model fit measures, including χ^2 ($df = 3$),[3] comparative fit index (CFI), root mean square error of approximation ($RMSEA$), and standardized root mean square residual ($SRMR$), were computed for the tested model. To test the index of moderated moderated mediation, a bias-corrected bootstrap confidence interval was constructed by 1,000 bootstrap samples. The estimates of a_1, a_3, b_1, and b_3 were retrieved from the tested model to transform W and Z to W' and Z'. Then, the correlations between (i) Z'^2 and $\frac{1}{W'^2}$, (ii) W'^2 and $\frac{1}{Z'^2}$, and (iii) the log-transformed W'^2 and Z'^2 were computed for evaluation.

4.3. Results of the Simulation Study

In the following sections, we would mainly discuss the results of the conditions where a spurious moderated mediation effect occurs. For these conditions, the simulation results on the model fit measures and model coefficients of the dual stage moderated mediation model are presented in Tables 2 and 3, respectively. Three dependent measures were included to evaluate the model coefficients, namely relative bias for point estimates, relative bias for standard error estimates, and empirical power. Table 4 presents two dependent measures to compare the inverse- and log-transformation methods, namely the average correlation across replications and the threshold of correlation based on percentile. Finally, for comparison purposes, we would briefly discuss the results of the conditions with a genuine moderated mediation effect. Due to space constraints, the simulation results of the comparison conditions are presented in supplementary materials (see Tables S1 and S2). In all tables, results are presented over the ten levels of sample size.

4.3.1. Model Fit Measures

To examine the goodness-of-fit of the dual stage moderated mediation model, we computed the average for model chi-square and three goodness-of-fit indices across the 1,000 replications (viz., $\overline{\chi^2}$, $\overline{CFI}$, $\overline{RMSEA}$, and $\overline{SRMR}$). Since the correct model is fitted to the data, the chi-square statistic is expected to follow a chi-square distribution with the corresponding degrees of freedom (i.e., $df = 3$). Therefore, the mean of the chi-square statistics across replications is expected to be 3. Based on the conventional cut-off criteria (Hu & Bentler, 1999), we consider a model having an acceptable fit to the data if $CFI > .95$, $RMSEA < .06$, and $SRMR < .06$. As shown in Table 2, the average chi-square statistic ($\overline{\chi^2}$) was close to its expected mean of 3 in all sample size conditions. All goodness-of-fit indices indicated a good fit to the data in all conditions; the model fit improved as the sample size increased.

[3]The covariance between ZM and e_M is a free parameter. The total number of free parameters is 25.

Table 2. Model Fit Measures across Conditions.

N	$\overline{\chi^2}$	$\overline{CFI}$	$\overline{RMSEA}$	$\overline{SRMR}$
100	3.139	0.996	0.034	0.020
200	3.034	0.998	0.022	0.014
300	3.109	0.999	0.019	0.012
400	3.110	0.999	0.016	0.011
500	3.081	0.999	0.015	0.009
600	3.214	0.999	0.014	0.009
700	3.004	0.999	0.012	0.008
800	3.095	1.000	0.012	0.008
900	2.934	1.000	0.010	0.007
1000	3.009	1.000	0.009	0.007

4.3.2. Model Coefficients

The relative percentage bias was used to evaluate the accuracy of the point and standard error estimates of the model coefficients (e.g., $\widehat{a_3}$ and $\widehat{b_3}$). They were defined as:

$$B\%\left(\hat{\beta}\right) = \frac{\overline{\hat{\beta}} - \beta}{\beta} \times 100,$$

$$B\%\left(\widehat{SE_{\hat{\beta}}}\right) = \frac{\overline{\widehat{SE_{\hat{\beta}}}} - ESD\left(\hat{\beta}\right)}{ESD\left(\hat{\beta}\right)} \times 100,$$

where β is the population value of a model coefficient, $\overline{\hat{\beta}}$ is the average of the point estimate of β across replications, $\overline{\widehat{SE_{\hat{\beta}}}}$ is the average of the standard error estimate of $\hat{\beta}$ across replications, and $ESD(\hat{\beta})$ is the empirical standard deviation of $\hat{\beta}$ across replications. A good estimation method should have a relative percentage bias of less than 10% for the point and standard error estimates (Flora & Curran, 2004; Hoogland & Boomsma, 1998). To evaluate the empirical power, the empirical rejection rate of $\hat{\beta}$ ($RR(\hat{\beta})$) at a level of significance of 5% was computed by dividing the number of significant replications by the total number of replications. A good estimation method should have an empirical power close to the nominal statistical power of 80%. The results on model coefficients are presented in Table 3.

4.3.2.1. Main Effects (a_1, a_2, b_1, b_2, and c). In all sample size conditions, the point and standard error estimates of all main effects showed a relative percentage bias of less than 10%; the empirical power of all main effects was larger than 80%.

4.3.2.2. Moderation Effects (a_3 and b_3). The point and standard error estimates of the two moderation effects showed a trivial bias (i.e., a relative percentage bias of less than 10%) in all conditions. The two moderation effects showed an acceptable empirical power if $N = 200$ or larger.

4.3.2.3. Index of Moderated Moderated Mediation (a_3b_3). In all sample size conditions, the index of moderated moderated mediation showed a relative percentage bias of less than 10% in the point estimate and in the bootstrap standard error estimate. Based on the bias-corrected bootstrap confidence interval, the test of a_3b_3 yielded an empirical power of at least 80% when $N = 200$. In general, its empirical power was enhanced as the sample size increased, though the enhancement stabilized from $N = 300$.

Table 3. Relative Percentage Biases for the Point and Standard Error Estimates and Empirical Rejection Rate of Model Coefficients across Conditions.

N	a_1	a_2	a_3	b_1	b_2	b_3	c	a_3b_3	$\frac{a_1}{a_3}$	$\frac{b_1}{b_3}$
$B\%(\hat{\beta})$										
100	−0.13	0.09	0.42	0.44	0.85	0.95	−2.39	−0.11	−5.65	20.48
200	−0.07	0.12	0.89	0.48	−0.15	0.12	−0.75	0.71	−13.20	6.96
300	−0.11	0.03	−0.73	0.43	0.13	−0.59	−0.65	−1.42	4.23	4.96
400	−0.24	−0.13	0.33	0.29	0.05	0.05	0.11	0.46	3.38	3.54
500	0.60	−0.04	0.77	0.87	0.18	−0.06	−0.54	0.73	2.20	3.21
600	0.36	−0.06	0.52	0.04	0.25	−0.27	0.12	0.30	1.97	2.24
700	−0.23	−0.36	−0.44	0.49	−0.12	−0.32	−0.35	−0.71	1.89	2.43
800	−0.45	−0.28	−0.39	0.35	−0.10	0.66	−0.07	0.27	1.55	0.90
900	−0.21	−0.06	−1.01	0.22	−0.06	−0.48	0.16	−1.43	2.06	1.93
1000	−0.35	−0.02	0.23	−0.46	−0.18	0.05	0.29	0.30	0.63	0.55
$B\%\left(\widehat{SE}_{\hat{\beta}}\right)$										
100	−0.11	−0.21	−0.86	−0.93	−0.69	−0.23	−0.25	0.32	–	–
200	−0.20	−0.12	−0.29	−0.45	−0.35	0.05	−0.20	0.14	–	–
300	0.09	−0.03	0.08	−0.11	−0.10	0.00	−0.19	0.11	–	–
400	−0.16	−0.10	−0.26	0.03	−0.19	−0.19	0.10	−0.02	–	–
500	0.02	0.02	0.10	−0.09	0.09	−0.03	−0.05	0.04	–	–
600	−0.19	0.10	−0.13	−0.02	−0.15	−0.07	−0.06	−0.01	–	–
700	−0.02	0.00	−0.01	−0.06	0.01	−0.07	−0.04	0.00	–	–
800	−0.19	−0.06	−0.17	−0.01	−0.08	−0.03	−0.13	−0.01	–	–
900	0.07	−0.09	0.02	0.03	−0.04	−0.04	−0.10	0.01	–	–
1000	0.02	0.09	−0.13	0.00	−0.01	−0.04	−0.04	−0.03	–	–
$RR(\hat{\beta}))$										
100	93.1	98.5	61.2	89.1	99.9	79.7	94.2	45.8	–	–
200	100.0	100.0	87.5	99.1	100.0	98.3	100.0	85.0	–	–
300	100.0	100.0	96.7	99.9	100.0	99.8	100.0	96.7	–	–
400	100.0	100.0	98.7	100.0	100.0	99.9	100.0	98.6	–	–
500	100.0	100.0	99.9	100.0	100.0	100.0	100.0	99.9	–	–
600	100.0	100.0	99.8	100.0	100.0	100.0	100.0	99.9	–	–
700	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	–	–
800	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	–	–
900	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	–	–
1000	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	–	–

Note. B% = relative percentage bias; $\hat{\beta}$= point estimates; $SE_{\hat{\beta}}$= standard error estimates; RR = empirical rejection rate.

Altogether, even though the moderated mediation effect is manipulated to be spurious in these conditions, the dual stage moderated mediation model still fitted the data well and the index of moderated moderated mediation still showed a high empirical rejection rate leading one to conclude the presence of a moderated mediation effect.

4.3.3. Two Methods to Rule out the Spurious Moderated Mediation Effect

To evaluate the two proposed methods in indicating the spurious moderated mediation effect, we computed the average correlations between Z'^2 and $\frac{1}{W'^2}$ ($\overline{r_{\text{inv }1}}$), W'^2 and $\frac{1}{Z'^2}$ ($\overline{r_{\text{inv }2}}$), and the log-transformed W'^2 and Z'^2 ($\overline{r_{\log}}$) across replications. $\overline{r_{\text{inv }1}}$ and $\overline{r_{\text{inv }2}}$ are expected to be +1.0, while $\overline{r_{\log}}$ is expected to be −1.0. To outline the cut-offs for the correlations, we computed the percentiles for the three correlations across the 1,000 replications. As Equation (13) is considered as the null hypothesis (H_0), the distributions of these correlations are generated under H_0. The recommended cut-off for the correlation is expected to minimize the risks of making a false conclusion under H_0 such that the correlation fails to indicate a spurious moderated mediation effect when it does actually exist. Following a nominal level of 5% in Type I error rate (or 95% in 1 − Type I error rate), a 5$^{\text{th}}$ percentile was computed for $r_{\text{inv }1}$ and $r_{\text{inv }2}$ to obtain a size of correlation where 95% of the replications are equal to or larger than that cut-off (cf_{inv}), while a 95$^{\text{th}}$ percentile was computed for $r_{\log}$ where 95% of the replications are equal to or smaller than that cut-off ($cf_{\log}$).

4.3.3.1. Log-Transformation Method.

As shown in the upper panel of Table 4, $\overline{r_{\log}}$ was negative in all sample size conditions, aligned with its population value of −1.0. The negative correlation also approached its population value along with the increase of sample size. For instance, when the sample size is large (e.g., $N = 1000$), $\overline{r_{\log}}$ equals −.97, yielding a fairly accurate estimate of $\rho_{\log}$ with a relative percentage bias of around −3%. As shown in the lower panel of Table 4, the cut-offs of $r_{\log}$ ($cf_{\log}$) were negative in all sample size conditions. Given a particular sample size condition, there were 5% of replications that yielded $r_{\log}$ being equal to or larger than the corresponding $cf_{\log}$. For instance, given $N = 200$, a correlation of −.71 could be seen as a cut-off and there were 5% of replications having $r_{\log}$ larger than −.71. In other words, 95% of replications yielded a $r_{\log}$ between −.71 to −1.0. The cut-offs approached −1.0 along with an increase of sample size and were stabilized at around −0.80 starting from $N = 300$. The frequency distribution of $r_{\log}$ in the condition of $N = 1000$ is included in the supplementary materials as an example (see Figure S1).

4.3.3.2. Inverse-Transformation Method.

Both $\overline{r_{\text{inv }1}}$ and $\overline{r_{\text{inv }2}}$ were positive in all sample size conditions, aligned with its population value of +1.0. The positive correlation also approached its population value along with the increase

Table 4. Average Correlations and Cut-offs for Inverse- and Log-Transformation Methods.

	Inverse-Transformation		*Log-Transformation*
N	$\overline{r_{inv\ 1}}$	$\overline{r_{inv\ 2}}$	$\overline{r_{log}}$
100	0.529	0.622	−0.793
200	0.611	0.724	−0.913
300	0.660	0.748	−0.945
400	0.642	0.752	−0.951
500	0.675	0.763	−0.962
600	0.678	0.771	−0.966
700	0.702	0.780	−0.971
800	0.699	0.790	−0.973
900	0.722	0.790	−0.977
1000	0.686	0.784	−0.974
N	$cf_{inv\ 1}$	$cf_{inv\ 2}$	cf_{log}
100	−0.044	−0.014	−0.256
200	0.005	0.056	−0.705
300	0.040	0.104	−0.820
400	0.037	0.116	−0.828
500	0.049	0.150	−0.854
600	0.059	0.119	−0.879
700	0.057	0.147	−0.882
800	0.062	0.149	−0.894
900	0.075	0.157	−0.915
1000	0.061	0.158	−0.899

Note. $r_{inv\ 1}$ = correlation between Z'^2 and $\frac{1}{W'^2}$; $r_{inv\ 2}$ = correlation between W'^2 and $\frac{1}{Z'^2}$; r_{log} = correlation between the log-transformed W'^2 and Z'^2; *cf* = cut-off values based on percentiles.

of sample size. $\overline{r_{inv\ 1}}$ was generally more biased than $\overline{r_{inv\ 2}}$. For instance, when the sample size is large (e.g., $N = 1000$), $\overline{r_{inv\ 1}}$ equals .69 with a relative percentage bias of around −31%, whereas $\overline{r_{inv\ 2}}$ equals .78 with a relative percentage bias of around −22%. Compared to $\overline{r_{log}}$, it is also apparent that $\overline{r_{inv\ 1}}$ and $\overline{r_{inv\ 2}}$ generally showed a large extent of bias. As shown in the lower panel of Table 4, the cut-offs of $r_{inv\ 1}$ and $r_{inv\ 2}$ ($cf_{inv\ 1}$ and $cf_{inv\ 2}$) differed largely from its population value of +1.0. For instance, $cf_{inv\ 1}$ and $cf_{inv\ 2}$ were +0.06 and +0.16 respectively when the sample size was large ($N = 1000$), whereas both cut-off values were negative when the sample size was small ($N = 100$). Overall, these results indicated that the inverse-transformation method did not perform well to identify the specific nonlinear relationship between Z' and W'. The frequency distributions of $r_{inv\ 1}$ and $r_{inv\ 2}$ in the condition of $N = 1000$ is included in the supplementary materials (see Figures S2 and S3).

4.3.4. Comparisons with the Conditions with a Genuine Moderated Mediation Effect

For the comparison conditions, the average chi-square statistic ($\overline{\chi^2}$) was close to its expected mean of 3 in all sample size conditions, while all goodness-of-fit indices indicated a good fit to the data in all sample size conditions (see Table S1). In all sample size conditions, the point and standard error estimates of the main effects, moderation effects, and index of moderated moderated mediation showed a relative percentage bias of less than 10% (see Table S2). Consistent with the conditions where a spurious moderated mediation effect occurs, a sample size of 300 is required to obtain an empirical power of at least 80% on all parameters (see Table S2). As expected, these results indicated that the dual stage moderated mediation model in the comparison conditions fitted the data well and the index of moderated moderated mediation showed a high empirical rejection rate to conclude the presence of a moderated mediation effect. As W and Z had a negative linear relationship in the comparison conditions, the population values of r_{log} and r_{inv} are positive and negative respectively. Consistent with our expectations, as shown in Table S1, $\overline{r_{log}}$ was positive in all sample size conditions, while $\overline{r_{inv\ 1}}$ and $\overline{r_{inv\ 2}}$ were negative in all sample size conditions. More importantly, the observed values of r_{log} in all sample size levels were much larger than −0.8 (i.e., the suggested cut-off derived from the lower panel of Table 4), indicating that a genuine moderated mediation effect can be concluded. Overall, the comparison of the spurious and genuine moderated mediation effect conditions provided evidence for the validity of Equation (13), in which r_{log} is largely deviated from the value of −1.0 when the moderated mediation effect is not spurious. It is noteworthy that the cut-offs for r_{log} and r_{inv} were not computed for the comparison conditions as the distributions of such correlations simply represent the situations under the alternative hypothesis (H_1) with a non-ignorable moderated mediation effect.

5. A Real Data Example

With this real data example, we aimed at illustrating the use of the proposed method to rule out a spurious moderated mediation effect. Our simulation study has shown that the log-transformation method is more preferred than the inverse-transformation method and hence, we focused primarily on the use of log-transformation method in this example. Based on the simulation results, the cut-offs of the log-transformation method stabilized at around −0.80 starting from $N = 300$, one can therefore consider a cut-off of −0.80 to check if a moderated mediation effect is likely to be spurious as long as $N \geq 300$ (see also our discussion section below). Simply put, if the correlation between two log-transformed moderators (r_{log}) is larger than −0.80, then a genuine moderated mediation effect can be concluded.

5.1. Background

The real data example was based on the data from Chiu et al. (2021). In their study, a total of 83 retail and security teams were included, yielding a total of 302 individuals. In this demonstration, we focused primarily on the individual-level analysis ($n = 302$); interested readers can refer to Chiu et al. (2021) for team-level and multi-level analyses. Aligned with their investigations, we examined whether the mediation effect of one's humility on experienced incivility through collective humility would be moderated by group humility diversity (moderated the first stage of the process) and group incivility differentiation (moderated the second stage of the process). This yields a dual stage moderated mediation model. The responses to all variables were anchored on a 7-point Likert scale. The reliability of all the measures was good, ranging from $\alpha = .93$ to $\alpha = .96$.

5.2. Results

We tested the dual stage moderated mediation model using *lavaan* (Rosseel, 2012). Results revealed that the model fitted the data well, χ^2 (3) = 5.299, p = .151, CFI = .999, $SRMR$ = .026, $RMSEA$ = .050. Consistent with Chiu et al. (2021), the index of moderated moderated mediation (a_3b_3) was significant, b = −0.23, 95% bootstrap CI [−0.52, −0.03], indicating that the mediation effect of one's humility on experienced incivility through collective humility was moderated by both group humility diversity and group incivility differentiation. Specifically, the mediation effect was only significant when both group humility diversity and group incivility differentiation were low (see Table S3 for all the model coefficients).

Finally, to check if the moderated mediation effect identified in the model is likely to be spurious, we employed the log-transformation method in two steps. In Step 1, we used the estimated model coefficients in the dual stage moderated mediation model (viz., a_1, a_3, b_1, and b_3 in Table S3) to compute $k' = \frac{a_1}{a_3}$ and $m' = \frac{b_1}{b_3}$. Then, the two moderators, group humility diversity (i.e., W) and group incivility differentiation (i.e., Z), were transformed through $W' = W + k'$ and $Z' = Z + m'$. In Step 2, we took the log-transformation for the square of the transformed moderators (i.e., $\log(W'^2)$ and $\log(Z'^2)$), and computed the correlation between them ($r_{\log}$). In this real data example, $k' = \frac{1.001}{-0.813} = -1.23$ and $m' = \frac{-0.472}{0.288} = -1.64$, and the log-transformed group humility diversity and group incivility differentiation yielded a correlation of 0.249, which is larger than −0.80. Given that this real data example had a sample size larger than 300, we conclude that the moderated mediation effect identified in the model is unlikely to be spurious. The *R*-script for employing the two-step log-transformation method is also provided in the supplementary materials.

6. Discussion

The use of moderated mediation models in social and behavioral sciences has been exponentially increasing in the past two decades, revealing its value to applied researchers. In a moderated mediation model, the effect of interest always lies in whether the size or direction of a mediation effect is conditional on the moderators, yielding a moderated mediation effect. In this research, we examined the possibility of concluding a moderated mediation effect which does not exist in a dual stage moderated mediation model, yielding a spurious moderated mediation effect.

With a spurious effect, research findings may become invalid and unreliable. Critically, given the widespread and emerging interest in a moderated mediation effect, the detrimental impact of concluding a spurious moderated mediation effect can be far-reaching since research findings might then not properly guide professionals in their practice and decision-making. In organizational psychology, for example, it has been shown that newcomers' self-perceived creativity caused their supervisors to give better evaluations of their creativity and hence task performance. The first stage of the mediation pathways was moderated by supervisors' trust in newcomers, while the second stage of the pathways was moderated by supervisors' support for authentic self-expression, yielding a dual stage moderated mediation effect (Dufour et al., 2020). Following these research findings, human resources professionals and managerial staff may devote resources to boost supervisors' trust in and support for authentic self-expression in newcomers, thereby maximizing the influence of newcomers' creativity on their performance. Nonetheless, if the dual stage moderated mediation effect of supervisors' trust and support is indeed spurious, these strategies may not work to promote task performance among newcomers, resulting in a potential waste of resources. Therefore, in the present research, we aimed at providing a solution for applied researchers to rule out the possibility of concluding a spurious moderated mediation effect.

In most situations, the index of moderated moderated mediation (Hayes, 2018) performs well in concluding a moderated mediation effect in a dual stage moderated mediation model. Nonetheless, as illustrated by the above theorems and example, this index may conclude a spurious moderated mediation effect when the two moderators hold a specific nonlinear relationship shown in Equation (13). To help rule out the existence of spurious results, we proposed two simple methods (viz., the inverse- and log-transformation methods) and tested their empirical performance in a simulation study. Statistically, the point estimate of $r_{\log}$ was less biased than that of r_{inv}. Even in a small sample, $r_{\log}$ managed to have a fairly accurate point estimate, while r_{inv} showed a considerable amount of bias even when the sample size was large. The cut-off value of $r_{\log}$ was close to its population value and started to stabilize from a relatively small sample size (i.e., $N = 300$), while the cut-off value of r_{inv} differed largely from its population values in all conditions, yielding a high probability of concluding false negative results. Practically, only one set of $r_{\log}$ is required in the log-transformation method, while two sets of r_{inv} are required in the inverse-transformation method. Thus, it is possible for the two sets of r_{inv} to indicate two inconsistent conclusions. These benefits, both statistical and practical, mean the log-transformation method is preferred over the inverse-transformation method. When one attempts to test the index of moderated moderated mediation, we recommend testing the correlation between the two log-transformed moderators as well. On the one hand, this test is simple and easy to implement. On the other hand, this test can be impactful as it helps avoid concluding a spurious effect which can be quite destructive to scientific advancement.

Based on the simulation results on $cf_{\log}$ (see Table 4), we have two recommendations for researchers. First, researchers can consider a particular cut-off based on the sample size in their studies. For instance, when $N = 500$, one may consider a negative correlation of −.85 as a cut-off. If the correlation falls below −.85 (e.g., between −.85 and −1.0), one should interpret the moderated mediation effect with caution due to the risk of a spurious result. Second, as the cut-offs stabilized at around −0.80 starting from $N = 300$,

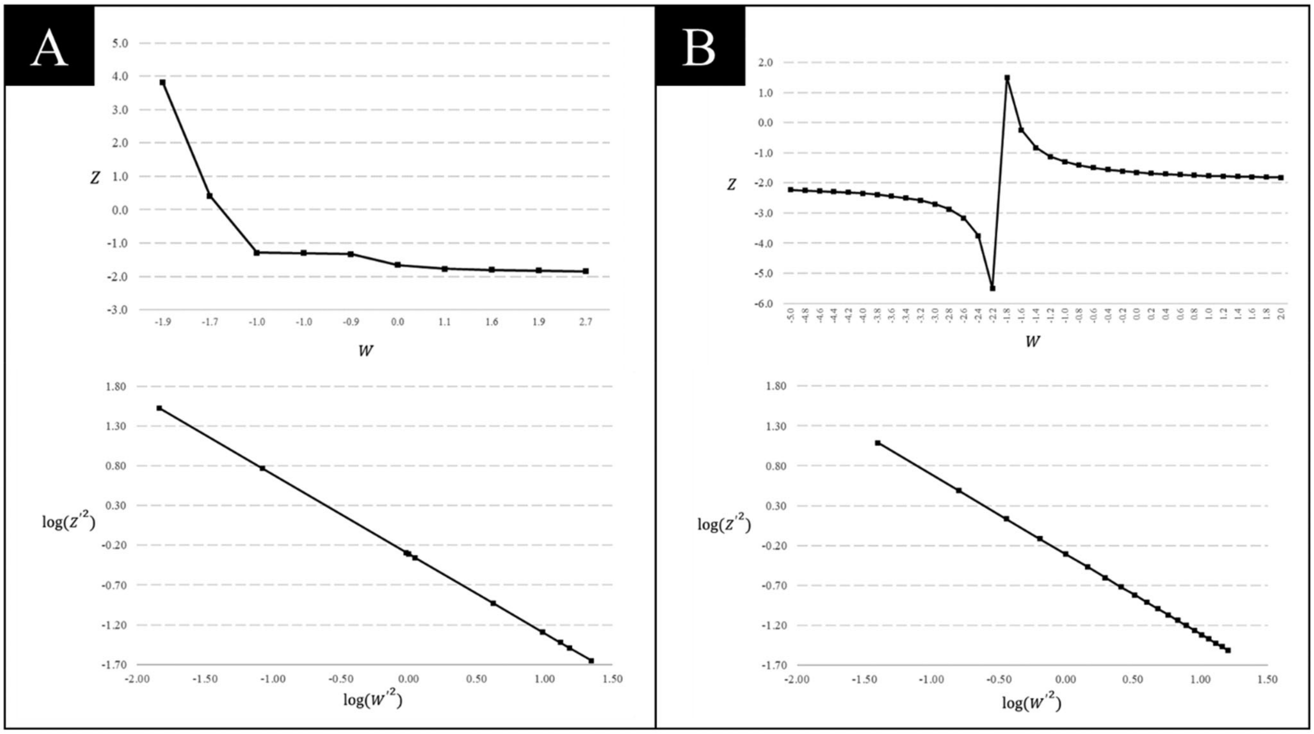

Figure 5. Scatterplots for the Relationships between W and Z and between $\log(W'^2)$ and $\log(Z'^2)$ in Two Data Examples.

one can take a more conservative approach by considering a cut-off of -0.80 as long as $N \geq 300$.[4] Nevertheless, the suggested test based on the log-transformation may not work reliably when the sample size is very small (i.e., $N = 100$) because it is likely for the cut-off to result in false negative results. To facilitate the use of the proposed methods, the R-script for the inverse- and log-transformation methods is provided in the supplementary materials.

Several points about the present investigation are worth noting. First, the performance of the two methods depends largely on an accurate transformation from W and Z to W' and Z'. If the estimates of k' (or $\frac{a_1}{a_3}$) and m' (or $\frac{b_1}{b_3}$) are less biased, the two methods should perform better. This expectation is confirmed by our simulation results in which the relative percentage bias of k' and m' (see the last two columns in the upper panel of Table 3) is directly proportional to the performance of both methods shown in Table 4. According to the simulation, the performance of the correlation between log-transformed moderators ($r_{\log}$) is better than the correlations based on the inverse-transformation method (r_{inv}). These findings show that the bias of k' and m' may differentially affect the moderators (e.g., W') and the inverse of the moderators (e.g., $\frac{1}{W'}$). In the current case, the inverse of the moderators was more susceptible to the bias of k' and m' than the moderators, thereby yielding a poorer performance in the inverse-transformation method. Future study is needed to examine under what circumstances the inverse-transformation method may perform better.

Second, the correlation between two moderators does not reveal whether a spurious moderated mediation effect occurs or not. The key reason for having a spurious moderated mediation effect is the existence of a specific relationship between the two moderators. Therefore, to identify the spurious effect, some researchers may think it is more straightforward to examine the correlation between two moderators instead of the three correlations in the two proposed methods. Consider the above hypothetical example, the correlation between W and Z is $-.64$ (see the visualized relationship in the upper panel of Figure 5A). Based on the same population model, we generated a larger sample of size $N = 35$, and the correlations between W and Z is $+.31$ (see the upper panel of Figure 5B). As shown in the upper panel of Figure 5, the patterns of the relationship between W and Z are quite different across the two data examples even though both are generated from the same population model. As a result, it is practically difficult to infer whether the condition in Equation (13) holds based on either the correlation between two moderators or the visualized relationship in scatterplot. Besides, as depicted in the nonlinear relationship in Figure 5B, it is possible for the correlation between two moderators to be positive, negative, or even null at the sample level, creating a huge difficulty to reveal the spurious effect based on that. In contrast, the methods proposed in this research are necessary and have the merit of indicating the spurious effect in both data examples. For instance, after we log-transformed the square of W' and Z'

[4]As the log-transformation method relies on the point estimates of $\frac{a_1}{a_3}$ and $\frac{b_1}{b_3}$, the cut-off values may vary depending on the size of the main and moderation effects. Thus, we also studied the cut-offs by varying the effect sizes of b_1 and b_3 (the effect sizes of a_1 and a_3 were kept constant for the sake of clearer comparison). In the current set of cut-offs, the effect sizes of b_1 and b_3 were relatively small (explained 7.85% and 9.54% of variance in Y). We studied two other conditions with (a) the relatively small size of b_1 (7.87%) and moderate size of b_3 (14.66%) and (b) the relatively moderate size of b_1 (13.16%) and small size of b_3 (7.69%). As shown in Table S4 in the supplementary materials, the cut-offs were comparable across the three conditions and stabilized at around $-.80$ starting from $N = 300$.

in both data examples, the correlation between $\log(W'^2)$ and $\log(Z'^2)$ is -1.00 in both examples (see the lower panel of Figure 5), indicating the log-ipsative property of W' and Z'.

Third, the current findings also have some implications for a dual stage moderated mediation model with a common moderator $\left(\begin{array}{ccccc} & & W & & \\ & \swarrow & & \searrow & \\ X & \rightarrow & M & \rightarrow & Y \end{array}\right)$. Given a common moderator W, the conditional effect of X on Y is

$$\gamma(W) = a_3 b_1 W + a_1 b_3 W + a_3 b_3 W^2. \tag{18}$$

For $\gamma(W) = \gamma$, the following equation has to be satisfied:

$$\begin{aligned} a_3 b_1 W + a_1 \mathrm{b}_3 W + a_3 b_3 W^2 &= \gamma \\ m' W + k' W + W^2 &= \gamma' \\ W^2 + (k' + m')W - \gamma' &= 0. \end{aligned} \tag{19}$$

Equation (19) is a quadratic equation so it can only have a maximum of two roots. In other words, when the moderator W has three or more distinctive values in the population, the conditional effect $\gamma(W)$ cannot be equal to a constant. In social and behavioral science research, however, it is common to have a binary moderator (e.g., gender). In this case, it is still possible for $\gamma(W) = \gamma$ to occur, yielding a spurious moderated mediation effect.[5]

Fourth, the present discussion can be extended to other models in which an effect varies as a multiplicative function of variables. For instance, consider a three-way interaction model $\left(\begin{array}{cccc} & W & & \\ & \downarrow & \leftarrow & Z \\ X & \rightarrow & Y & \end{array}\right)$:

$$Y = a_0 + a_1 X + a_2 W + a_3 Z + a_4 XW + a_5 XZ + a_6 WZ + a_7 XWZ + e_Y,$$

the effect of X on Y is conditional on the multiplication of W and Z and can be written as $a_1 + a_4 W + a_5 Z + a_7 WZ$. Thus, the unconditional effect of X on Y is a_1 while the conditional effect is $a_4 W + a_5 Z + a_7 WZ$. The conditional effect in a three-way interaction model is essentially the same as the conditional effect in Equation (4). Hence, the three theorems presented above also apply to a three-way interaction model, indicating a possibility of concluding a spurious three-way interaction effect. Following the same rationale, the present work applies to other complex moderated mediation models where a three-way interaction is involved either on the first stage (Model 11 in PROCESS) or second stage (Model 18 in PROCESS) of a mediation process $\left(\begin{array}{ccccc} & & W & & \\ Z & \rightarrow & \downarrow & & \\ & X & \rightarrow & M & \rightarrow & Y \end{array}\right.$ and $\left.\begin{array}{ccccc} & & & W & \\ & & & \downarrow & \leftarrow & Z \\ X & \rightarrow & M & \rightarrow & Y & \end{array}\right)$.

6.2. Limitations and Future Directions

There are several limitations in our simulation study. The current simulation study assumed that the variables are measurement error-free. Since the unreliability of variables will lead to a biased estimation of correlations, future research should take the unreliability issue into consideration and examine ways to account for measurement errors in the proposed methods (e.g., latent variable, factor score, and reliability-correction approaches). Besides, we only studied a limited number of configurations of model parameters for the dual stage moderated mediation model in the current study. Although the choice of the parameter values was aligned with the common sizes reported in previous studies (see also Footnote 4), a wider range of effect sizes should be systematically manipulated to study how the sizes of main and moderation effects affect the performance of the two methods.

In the current simulation study, the moderators were generated from an F distribution because the inverse of F distribution has well-known distribution properties, thereby making the data generation process much simpler. Specifically, during the data generation process, the mean and variance of Z in Equation (13) can be well-defined because the inverse of W is well-defined in F distribution. Given the current simulation settings, it is still unclear whether the proposed methods will perform well when the moderators are generated from other distributions. Accordingly, the cut-offs may also vary depending on the distributions of the moderators. For this purpose, an additional simulation study was conducted to glance over the performance of the log-transformation method when moderators are generated from two other distributions (viz., a uniform and a truncated normal distributions).[6] Results revealed that the performance of the log-transformation method in these two distributional conditions was generally comparable to the main simulation study (see Table S5). Simply put, the average correlations ($\overline{r_{\log}}$) were negative in all sample size conditions and approached its population value of -1.0 along with the increase of sample size, while the cut-offs ($cf_{\log}$) were also stabilized at around -0.80 starting from $N = 300$ (the cut-offs in a uniform distribution were even stabilized at around -0.90). Overall, our recommendation of considering a cut-off of -0.80 as long as

[5]Let W be a binary variable with two possible values, 0 and $-(k' + m')$. Here, in addition to the assumptions $a_3 \neq 0$ and $b_3 \neq 0$, we have to further assume that $k' \neq -m'$ (or $a_1 b_3 \neq -a_3 b_1$) to ensure that $-(k' + m')$ is a distinctive value other than zero. Since the coding of a binary variable is basically arbitrary, assigning 0 and $-(k' + m')$ as the codes of W is still a legitimate practice. For example, if $k' = m' = -0.5$, then $-(k' + m') = 1$. Following Equation (18), we therefore have

$$\gamma(0) = a_3 b_1 (0) + a_1 b_3 (0) + a_3 b_3 (0^2) = 0$$

$$\gamma(-(k'+m')) = a_3 b_1 (-(k'+m')) + a_1 b_3 (-(k'+m')) + a_3 b_3 (-(k'+m'))^2$$

$$= -a_1 b_1 - \frac{a_3 b_1^2}{b_3} - \frac{a_1^2 b_3}{a_3} - a_1 b_1 + \frac{a_1^2 b_3}{a_3} + \frac{a_3 b_1^2}{b_3} + 2 a_1 b_1 = 0$$

Consequently, $\gamma(W) = 0$ for both values of W.

[6]In social sciences research, variables are typically measured in an interval scale (e.g., a 5-point Likert scale). To approximate the data in social sciences research, we further examined two distributional conditions of W. Specifically, W was drawn from 1) a uniform distribution on an interval between 1 and 5 and 2) a truncated normal distribution on an interval between 1 and 5 with a mean of 3 and unit variance. Other than the distribution of the moderators, all the simulation settings are the same as those in the main simulation study.

$N \geq 300$ is also applicable to these two distributional conditions. Further research should be conducted to evaluate the proposed methods in other distributions (e.g., the χ^2 distributions that reflect various degrees of skewness and kurtosis).

Finally, apart from the methods proposed in this paper, future research should explore other methods to check if the spurious moderated mediation effect occurs. A possible alternative is to directly estimate the nonlinear relationship between the two moderators of W and Z as in the form of Equation (13). Then, one can compare the estimated m' and k' with those obtained from testing a dual stage moderated mediation model. This comparison may provide insights on whether the specific non-linear relationship between W and Z exists, yielding a spurious moderated mediation effect. Yet, this approach is more complicated that the log-transformation method proposed in this paper since it requires (1) fitting a nonlinear relationship between W and Z as well as (2) testing the significance of differences between two sets of estimated m' and k'. In any case, further research is needed to examine the empirical performance of this method.

In conclusion, this article has illustrated the possibility of concluding a spurious moderated mediation effect and proposed two methods to rule out this possibility. Simulation results have confirmed the usefulness of the methods and helped to outline some cut-off values for the applied researcher to use to check if their moderated mediation effects are likely to be spurious. Potential extensions of the present work to other statistical models common in social and behavioral sciences were also discussed. Based on the present investigation, we would recommend the use of the log-transformation method to rule out the possibility of having spurious results when the index of moderated moderated mediation is tested. Given that it is uncertain whether this spurious issue is common or not in applied research, we strongly recommend researchers to check before a firm conclusion is made. More importantly, even though this phenomenon is not frequent, the benefits of the test should outweigh its costs since the test is simple and easy to implement while its impact of ensuring valid findings can be far-reaching.

References

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. *The Journal of Applied Psychology*, *90*, 94–107. https://doi.org/10.1037/0021-9010.90.1.94

Aguinis, H., Edwards, J. R., & Bradley, K. J. (2017). Improving our understanding of moderation and mediation in strategic management research. *Organizational Research Methods*, *20*, 665–685. https://doi.org/10.1177/1094428115627498

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

Armstrong, B., Carmody, J. K., & Janicke, D. M. (2014). Predictors of maternal encouragement to diet: A moderated mediation analysis. *Maternal and Child Health Journal*, *18*, 1480–1487. https://doi.org/10.1007/s10995-013-1388-5

Bamberger, P., & Belogolovsky, E. (2017). The dark side of transparency: How and when pay administration practices affect employee helping. *The Journal of Applied Psychology*, *102*, 658–671. https://doi.org/10.1037/apl0000184

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173–1182. https://doi.org/10.1037/0022-3514.51.6.1173

Bauer, D. J., Preacher, K. J., & Gil, K. M. (2006). Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: New procedures and recommendations. *Psychological Methods*, *11*, 142–163. https://doi.org/10.1037/1082-989x.11.2.142

Chiu, C. Y. C., Marrone, J. A., & Tuckey, M. R. (2021). How do humble people mitigate group incivility? An examination of the social oil hypothesis of collective humility. *Journal of Occupational Health Psychology*, *26*, 361–373. https://doi.org/10.1037/ocp0000244

Champoux, J. E., & Peters, W. S. (1987). Form, effect size and power in moderated regression analysis. *Journal of Occupational Psychology*, *60*, 243–255. https://doi.org/10.1111/j.2044-8325.1987.tb00257.x

Chaplin, W. F. (1991). The next generation of moderator research in personality psychology. *Journal of Personality*, *59*, 143–178. https://doi.org/10.1111/j.1467-6494.1991.tb00772.x

Chen, S. P., & Cheng, C. P. (2014). Model specification for latent interactive and quadratic effects in matrix form. *Structural Equation Modeling*, *21*, 94–101. https://doi.org/10.1080/10705511.2014.859509

Cheung, G. W., & Lau, R. S. (2017). Accuracy of parameter estimates and confidence intervals in moderated mediation models. *Organizational Research Methods*, *20*, 746–769. https://doi.org/10.1177/1094428115595869

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. Routledge.

Dufour, L., Maoret, M., & Montani, F. (2020). Coupling high self-perceived creativity and successful newcomer adjustment in organizations: The role of supervisor trust and support for authentic self-expression. *Journal of Management Studies*, *57*, 1531–1555. https://doi.org/10.1111/joms.12547

Edwards, J. R., & Lambert, L. S. (2007). Methods for integrating moderation and mediation: A general analytical framework using moderated path analysis. *Psychological Methods*, *12*, 1–22. https://doi.org/10.1037/1082-989x.12.1.1

Fairchild, A. J., & MacKinnon, D. P. (2009). A general model for testing mediation and moderation effects. *Prevention Science*, *10*, 87–99. https://doi.org/10.1007/s11121-008-0109-6

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, *9*, 466–491. https://doi.org/10.1037/1082-989x.9.4.466

Foldnes, N., & Hagtvet, K. A. (2014). The choice of product indicators in latent variable interaction models: Post hoc analyses. *Psychological Methods*, *19*, 444–457. https://doi.org/10.1037/a0035728

Gopalan, N., Pattusamy, M., & Goodman, S. (2022). Family incivility and work-engagement: Moderated mediation model of personal resources and family-work enrichment. *Current Psychology*, *41*, 7350–7361. https://doi.org/10.1007/s12144-021-01420-4

Harring, J. R., Weiss, B. A., & Li, M. (2015). Assessing spurious interaction effects in structural equation modeling. *Educational and Psychological Measurement*, *75*, 721–738. https://doi.org/10.1177/0013164414565007

Hayes, A. F. (2015). An index and test of linear moderated mediation. *Multivariate Behavioral Research*, *50*, 1–22. https://doi.org/10.1080/00273171.2014.962683

Hayes, A. F. (2018). Partial, conditional, and moderated moderated mediation: Quantification, inference, and interpretation. *Communication Monographs*, *85*, 4–40. https://doi.org/10.1080/03637751.2017.1352100

Hayes, A. F. (2022). *Introduction to mediation, moderation, and conditional process analysis* (3rd ed.). The Guilford Press.

Hayes, A. F., Montoya, A. K., & Rockwood, N. J. (2017). The analysis of mechanisms and their contingencies: Process versus structural equation modeling. *Australasian Marketing Journal*, *25*, 76–81. https://doi.org/10.1016/j.ausmj.2017.02.001

Hayes, A. F., & Rockwood, N. J. (2017). Regression-based statistical mediation and moderation analysis in clinical research: Observations, recommendations, and implementation. *Behaviour Research and Therapy*, *98*, 39–57. https://doi.org/10.1016/j.brat.2016.11.001

Hayes, A. F., & Rockwood, N. J. (2020). Conditional process analysis: Concepts, computation, and advances in the modeling of the contingencies of mechanisms. *American Behavioral Scientist*, *64*, 19–54. https://doi.org/10.1177/0002764219859633

Hoogland, J. J., & Boomsma, A. (1998). Robustness studies in covariance structure modeling. *Sociological Methods & Research*, *26*, 329–367. https://doi.org/10.1177/0049124198026003003

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–55. https://doi.org/10.1080/10705519909540118

Igartua, J. J., & Hayes, A. F. (2021). Mediation, moderation, and conditional process analysis: Concepts, computations, and some common confusions. *The Spanish Journal of Psychology*, *24*, 1–23. https://doi.org/10.1017/SJP.2021.46

Jaccard, J., & Wan, C. K. (1995). Measurement error in the analysis of interaction effects between continuous predictors using multiple regression: Multiple indicator and structural equation approaches. *Psychological Bulletin*, *117*, 348–357. https://doi.org/10.1037/0033-2909.117.2.348

James, L. R., & Brett, J. M. (1984). Mediators, moderators, and tests for mediation. *Journal of Applied Psychology*, *69*, 307–321. https://doi.org/10.1037/0021-9010.69.2.307

Kim, S., & Hong, S. (2020). Comparing methods for multilevel moderated mediation: A decomposed-first strategy. *Structural Equation Modeling*, *27*, 661–677. https://doi.org/10.1080/10705511.2019.1683015

Klein, A., & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrika*, *65*, 457–474. https://doi.org/10.1007/BF02296338

Kromrey, J. D., & Foster-Johnson, L. (1999). Statistically differentiating between interaction and Nonlinearity in multiple regression analysis: A Monte Carlo investigation of a recommended strategy. *Educational and Psychological Measurement*, *59*, 392–413. https://doi.org/10.1177/00131649921969947

Kwan, J. L. Y., & Chan, W. (2018). Variable system: An alternative approach for the analysis of mediated moderation. *Psychological Methods*, *23*, 262–277. https://doi.org/10.1037/met0000160

Laran, J., Dalton, A. N., & Andrade, E. B. (2011). The curious case of behavioral backlash: Why brands produce priming effects and slogans produce reverse priming effects. *Journal of Consumer Research*, *37*, 999–1014. https://doi.org/10.1086/656577

Lemmer, G., & Gollwitzer, M. (2017). The "true" indirect effect won't (always) stand up: When and why reverse mediation testing fails. *Journal of Experimental Social Psychology*, *69*, 144–149. https://doi.org/10.1016/j.jesp.2016.05.002

Li, A., Shaffer, J., & Bagger, J. (2015). The psychological well-being of disability caregivers: Examining the roles of family strain, family-to-work conflict, and perceived supervisor support. *Journal of Occupational Health Psychology*, *20*, 40–49. https://doi.org/10.1037/a0037878

Little, T. D., Card, N. A., Bovaird, J. A., Preacher, K. J., & Crandall, C. S. (2007). Structural equation modeling of mediation and moderation with contextual factors. In T. D. Little, J. A. Bovaird & N. A. Card (Eds.), *Modeling contextual effects in longitudinal studies* (pp. 207–230). Lawrence Erlbaum Associates.

Liu, H., Yuan, K. H., & Wen, Z. (2022). Two-level moderated mediation models with single-level data and new measures of effect sizes. *Behavior Research Methods*, *54*, 574–596. https://doi.org/10.3758/s13428-021-01578-6

Loeys, T., Talloen, W., Goubert, L., Moerkerke, B., & Vansteelandt, S. (2016). Assessing moderated mediation in linear models requires fewer confounding assumptions than assessing mediation. *The British Journal of Mathematical and Statistical Psychology*, *69*, 352–374. https://doi.org/10.1111/bmsp.12077

Lubinski, D., & Humphreys, L. G. (1990). Assessing spurious "moderator effects": Illustrated substantively with the hypothesized ("synergistic") relation between spatial and mathematical ability. *Psychological Bulletin*, *107*, 385–393. https://doi.org/10.1037/0033-2909.107.3.385

MacKinnon, D. P., & Fairchild, A. J. (2009). Current directions in mediation analysis. *Current Directions in Psychological Science*, *18*, 16–20. https://doi.org/10.1111/j.1467-8721.2009.01598.x

Marsh, H. W., Wen, Z., & Hau, K. T. (2004). Structural equation models of latent interactions: Evaluation of alternative estimation strategies and indicator construction. *Psychological Methods*, *9*, 275–300. https://doi.org/10.1037/1082-989x.9.3.275

Maxwell, S. E., & Delaney, H. D. (1993). Bivariate median splits and spurious statistical significance. *Psychological Bulletin*, *113*, 181–190. https://doi.org/10.1037/0033-2909.113.1.181

Mitchell, M. S., Vogel, R. M., & Folger, R. (2015). Third parties' reactions to the abusive supervision of coworkers. *The Journal of Applied Psychology*, *100*, 1040–1055. https://doi.org/10.1037/apl0000002

Moulder, B. C., & Algina, J. (2002). Comparison of methods for estimating and testing latent variable interactions. *Structural Equation Modeling*, *9*, 1–19. https://doi.org/10.1207/S15328007SEM0901_1

Muller, D., Judd, C. M., & Yzerbyt, V. Y. (2005). When moderation is mediated and mediation is moderated. *Journal of Personality and Social Psychology*, *89*, 852–863. https://doi.org/10.1037/0022-3514.89.6.852

Nelder, J. A. (1977). A reformulation of linear models. *Journal of the Royal Statistical Society. Series A (General)*, *140*, 48–77. https://doi.org/10.2307/2344517

Ng, J. C. K., & Chan, W. (2020). Latent moderation analysis: A factor score approach. *Structural Equation Modeling*, *27*, 629–648. https://doi.org/10.1080/10705511.2019.1664304

Ng, J. C. K., Chan, W., Kwan, J. L. Y., & Chen, S. X. (2019). Unpacking structure-oriented cultural differences through a mediated moderation model: A tutorial with an empirical illustration. *Journal of Cross-Cultural Psychology*, *50*, 358–380. https://doi.org/10.1177/0022022118821183

Preacher, K. J. (2015). Advances in mediation analysis: A survey and synthesis of new developments. *Annual Review of Psychology*, *66*, 825–852. https://doi.org/10.1146/annurev-psych-010814-015258

Preacher, K. J., Rucker, D. D., & Hayes, A. F. (2007). Addressing moderated mediation hypotheses: Theory, methods, and prescriptions. *Multivariate Behavioral Research*, *42*, 185–227. https://doi.org/10.1080/00273170701341316

Racine, S. E., & Martin, S. J. (2017). Integrating eating disorder-specific risk factors into the acquired preparedness model of dysregulated eating: A moderated mediation analysis. *Eating Behaviors*, *24*, 54–60. https://doi.org/10.1016/j.eatbeh.2016.12.007

R Development Core Team. (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.

Rosseel, Y. (2012). lavaan: An *R* Package for structural equation modeling. *Journal of Statistical Software*, *48*, 1–36. https://doi.org/10.18637/jss.v048.i02

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods*, *7*, 422–445. https://doi.org/10.1037/1082-989X.7.4.422

Simon, H. A. (1954). Spurious correlation: A causal interpretation. *Journal of the American Statistical Association*, *49*, 467–479. https://doi.org/10.2307/2281124

Thoemmes, F., MacKinnon, D. P., & Reiser, M. R. (2010). Power analysis for complex Mediational designs using Monte Carlo methods. *Structural Equation Modeling*, *17*, 510–534. https://doi.org/10.1080/10705511.2010.489379

Wang, L., & Preacher, K. J. (2015). Moderated mediation analysis using Bayesian methods. *Structural Equation Modeling*, *22*, 249–263. https://doi.org/10.1080/10705511.2014.935256

Wedel, M., & Dong, C. (2022). A tutorial on the analysis of experiments using BANOVA. *Psychological Methods*, *27*, 433–450. https://doi.org/10.1037/met0000408

Yaffe, T., & Kark, R. (2011). Leading by example: The case of leader OCB. *The Journal of Applied Psychology*, *96*, 806–826. https://doi.org/10.1037/a0022464

Yuan, K. H., & Gomer, B. (2021). An overview of applied robust methods. *British Journal of Mathematical and Statistical Psychology*, *74*, 199–246. https://doi.org/10.1111/bmsp.12230

Yzerbyt, V., Muller, D., Batailler, C., & Judd, C. M. (2018). New recommendations for testing indirect effects in mediational models: The need to report and test component paths. *Journal of Personality and Social Psychology*, *115*, 929–943. https://doi.org/10.1037/pspa0000132