A Robust Bootstrap Test for Mediation Analysis: Online Appendix

Andreas Alfons

alfons@ese.eur.nl +3110 408 2588

Econometric Institute, Erasmus University Rotterdam

Burgemeester Oudlaan 50, 3062 PA Rotterdam, Netherlands

Nüfer Yasin Ates

nufer.ates@sabanciuniv.edu +90 216 483 90 00

Sabancı Business School, Sabancı University

Orta Mahalle, 34956 Tuzla, Istanbul, Turkey

Patrick J.F. Groenen

groenen@ese.eur.nl +3110 408 1281

Econometric Institute, Erasmus University Rotterdam

Burgemeester Oudlaan 50, 3062 PA Rotterdam, Netherlands

A Robust Bootstrap Test for Mediation Analysis: Online Appendix

Online Appendix 1: Additional Information on the Illustrative Empirical Case

Data. The data for the illustrative case come from a larger research program on team processes. Data were collected from 354 senior business administration students playing a 12-round business simulation game (two separate games of 6 rounds) in randomly assigned 4-person teams (92 teams in total) as part of their capstone strategy course at a Western European university. Data on several individual- and team-level constructs were collected in three survey waves: prior to, during, and after the simulation game, with different constructs being surveyed in the different waves. The overall response rate was 93% (332 students). Leaving out teams with less than 50% response rate yields n = 89 teams for further analysis.

Theory. Values are standards that guide thought and action (Schwartz, 1992). Values predispose individuals to favor one ideology over another, determine how one judges oneself and others, and cause taking certain positions on social issues (Rokeach, 1973). Schwartz's value theory proposes ten distinct universal values that are theoretically derived from human nature; these ten values are power, achievement, hedonism, stimulation, self-direction, universalism, benevolence, tradition, conformity, and security. When team members possess different set of values – meaning a high team value diversity – teams can experience higher levels of conflict in executing their tasks (Jehn, 1994), because the variety of worldviews may cause different prioritizations of actions that need to be coherently conducted. Conflict on the task content triggered by a difference in values can be detrimental to team outcomes (Jehn, Northcraft, & Neale, 1999), such as team commitment (Mowday, Steers, & Porter, 1979). Team commitment is the strength of team members' identification with, and involvement in, a particular team (Bishop & Scott, 2000). Accordingly, we propose that team value diversity affects team commitment through task conflict.

Measures, validity and reliability. We operationalized *task conflict* with the intra-group task conflict scale of (Jehn, 1995). The five items on the presence of conflict were rated on a 5-point Likert scale (anchored by 1 = “None” and 5 = “A lot”). Sample items measuring task conflict include the following: “How frequently are there conflicts about ideas in your work unit?” and “How often do people in your work unit disagree about opinions?”. Cronbach’s alpha is 0.86. We aggregated individual responses to team level (median $r_{WG} = 0.95$). We used the short version of Schwartz’s Value Survey (SVS) to measure team members’ individual values (Lindeman & Verkasalo, 2005). Then we operationalized *value diversity* with average of the coefficient of variation of each value dimension among team members. *Team commitment* is measured by four items based on Mowday, Steers, & Porter (1979). Sample items include “I feel proud to belong to this team” and “I am willing to exert extra effort to help this team succeed”. Cronbach’s alpha is 0.78. Individual responses were aggregated to team level (median $r_{WG} = 0.93$). Value diversity was measured in survey 1, task conflict in survey 2, and team commitment in survey 3.

Online Appendix 2: Additional Simulation Studies

In addition to the simulation studies reported in the manuscript, we ran several alternative simulations as robustness checks. These simulations cover a wide range of settings to attribute further reliability to our results. We compare the following methods:[1]

- *OLS bootstrap:* the bootstrap test following OLS estimation (Bollen & Stine, 1990; Shrout & Bolger, 2002; MacKinnon, Lockwood, & Williams, 2004; Preacher & Hayes, 2004; Preacher & Hayes, 2008);
- *OLS Sobel:* the Sobel test following OLS estimation (Sobel, 1982);
- *Box-Cox bootstrap:* we first apply Hawkins & Weisberg's (2017) generalization of the Box-Cox transformation (Box & Cox, 1964) to each variable, then perform the OLS bootstrap test;
- *Winsorized bootstrap:* Zu & Yuan's (2010) bootstrap test following winsorization of the data;
- *Median bootstrap:* Yuan & MacKinnon's (2014) bootstrap test using median regression;
- *ROBMED:* our proposed test using MM-estimation (Yohai, 1987) and the fast-and-robust bootstrap (Salibián-Barrera & Zamar, 2002).

All bootstrap tests use $R = 5000$ bootstrap samples, and report a bias-corrected and accelerated percentile-based confidence interval (Davison & Hinkley, 1997) for the indirect effect. The data are generated according to the models $M = aX + \sigma_1 e_1$ and $Y = bM + cX + \sigma_2 e_2$. We thereby use four different simulation designs:

1. The first simulation design is a variation of the design from the manuscript, but with varying sample sizes n and varying effect sizes of a, b, and c. The

[1] We exclude the SNT bootstrap (see Table 3 of the manuscript) from the additional simulations due to its long computation time.

aim of this design is to investigate how robust the findings of the manuscript are across a range of typical sample sizes and effect sizes in organizational research.

2. The second simulation design uses the same basic design from the manuscript, but with varying probability of outliers and varying distance of the outliers from the main part of the data. The aim of this design is to investigate how the methods react to different outlier settings.
3. The third simulation design uses the same basic design from the manuscript, but uses centered log-normal distributions, skew-t distributions, and distributions generated by Fleishman's method (Fleishman, 1978). Various parameter settings are used in the latter two cases. The aim of this design is to investigate how the methods are affected by different levels of skewness and kurtosis.
4. The fourth simulation design consists of extensions of Zu & Yuan's (2010) design with varying percentage of outliers and varying distance of outliers from the main part of the data. The aim of this design is to verify our findings regarding outliers.
5. The fifth simulation design is taken from Yuan & MacKinnon (2014) with different distributions of the error terms. This design also varies the number of observations and the effect sizes. Its aim is to verify our findings regarding different error distributions.

For each simulation design, we compare the methods in two situations: (i) when there is mediation, and (ii) when there is no mediation. In total, this yields 700 different parameter settings, which allows us to draw robust conclusions about the performance of the methods.

We generate $K = 1000$ data sets for each setting. On each data set, two-sided tests with null hypothesis $H_0: ab = 0$ against the alternative $H_a: ab \neq 0$ are performed.

Note that the Box-Cox bootstrap applies a nonlinear transformation to each variable, therefore the estimates are not comparable to the other methods (in particular under deviations from the model assumptions). This already highlights a disadvantage of nonlinear transformations, as in this case we no longer get an estimate of the actual model parameter, but an estimate for a different model that is difficult to interpret (see also Becker, Robertson, & Vandenberg, 2019). We therefore discuss the performance of the Box-Cox bootstrap mainly in terms of the significance test of the indirect effect, and in general exclude it from discussions on the accuracy of the estimates of the indirect effect.

Simulation design 1: Robustness of findings across range of sample sizes and effect sizes

The explanatory variable X is generated from a standard normal distribution. We explore a range of effect sizes in two different situations: one with mediation ($a = b = c = 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8$, true indirect effect $ab = 0.04, 0.09, 0.16, 0.25, 0.36, 0.49, 0.64$), and one where mediation does not exist ($a = c = 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8$, $b = 0$, true indirect effect $ab = 0$). We consider the same four settings regarding the error distributions and outliers as described in Table 5 in the manuscript. The parameter σ_1 is chosen such that M has variance 1 in the setting with normally distributed errors, but unlike in the manuscript, we always set $\sigma_2 = 1$.[2]

Varying effect size. We first give a detailed discussion of the results for varying effect size for sample size $n = 100$, which is the sample size used in the manuscript. The conclusions are similar for the other sample sizes, except that the power of all methods

[2] With this data generating process, the variance of Y is given by $\sigma_Y^2 = b^2 + c^2 + abc + \sigma_2^2$. Hence for large enough values of a, b and c, it is no longer possible to restrict the variance of Y to 1. We compensate for this by adjusting the values of outliers to $M_i^* = M_i/10 - 3$ and $Y_i^* = Y_i/10 + 3\sigma_Y$.

increases with increasing sample size. The results for varying effect size in different sample sizes are shown in Figures 1–10.

Simulations with mediation. Figure 2 displays the results for sample size $n = 100$. The top row shows the average bias of the indirect effect (that is, the average of $\widehat{ab} - ab$ over the simulation runs) for increasing effect size, while the bottom row shows the rate of how often the methods reject the null hypothesis and the corresponding estimate of the indirect effect has the correct sign. Note that evaluating the methods by the rejection rate from the two-sided tests alone does not provide a meaningful comparison in these simulation settings, as outliers and other deviations from normality can push the estimated indirect effect from a positive one towards a negative one. This incorrectly estimated negative indirect effect can be large enough in magnitude to reject the null hypothesis of a two-sided test. However, while the sign of the estimated effect is negative, the sign of the true effect is positive, which would result in an incorrect interpretation of the indirect effect. By taking into account the sign of the estimated indirect effect as well, we obtain a better measure of realized power of the tests.

With normal error terms, all methods estimate the indirect effect very accurately, although it seems that the Box-Cox bootstrap has an increasing tendency to apply unnecessary transformations as the effect sizes increase. For small effect sizes, ROBMED is slightly less powerful than the OLS bootstrap, the winsorized bootstrap, or the Box-Cox bootstrap, but it does not lose much power. Furthermore, it is more powerful than the OLS Sobel test and the median bootstrap.

In the presence of outliers, ROBMED is the only method that still gives accurate estimates of the indirect effect. The OLS-based methods (i.e., OLS bootstrap and OLS Sobel) are the most affected by the outliers, while the median bootstrap and the winsorized bootstrap also show a considerable bias. For all of these methods, the bias increases with increasing effect size. The results from estimation carry over to the realized power of the tests, with

ROBMED being clearly the most powerful test. The winsorized bootstrap is the only other method that is not too far behind in terms of power, despite its bias in effect size. Interestingly, the power of all methods except ROBMED first rises with increasing effect size, but then decreases again for large effect sizes. A possible explanation is that different effect sizes change the relative position of the outliers with respect to main data cloud, as the shape of that data cloud is changed. This means that the influence of outliers on the methods could be somewhat different for different effect sizes.

For skew-normal error terms, all methods are very accurate in estimating the indirect effect. The Box-Cox bootstrap, the winsorized bootstrap, the OLS bootstrap, and ROBMED are highly similar in terms of power, with only minor differences for the smallest investigated effect size. As in the setting with normal error terms, the median bootstrap has slightly less power than the other methods.

For t-distributed errors, all methods estimate the indirect effect accurately, although it should be noted that the Box-Cox bootstrap is deceived into applying power transformations that are not particularly suitable for symmetric heavy tails. In terms of power, ROBMED, the winsorized bootstrap, the median bootstrap, and the Box-Cox bootstrap all have high power, with ROBMED being marginally more powerful for small and moderate effect sizes. Only the OLS-based tests show a considerable loss of power.

Simulations with no mediation. Figure 7 shows the results for sample size $n = 100$. The top row of Figure 7 shows the average relative bias of the indirect effect with respect to the effect size of a. Note that while $b = 0$, estimates will be small but nonzero. When multiplying this nonzero estimate of b with the estimate of a, the resulting estimate of ab will naturally increase with increasing effect size a, and therefore also the bias compared to the true value $ab = 0$. The same holds of course when the estimate $\widehat{ab}$ is obtained as the average over the bootstrap replicates. Therefore, the relative bias $(\widehat{ab} - ab)/a$ is a more meaningful evaluation of the estimates of the indirect effect across different effect sizes of a. The bottom row of Figure 7 presents the rejection rate. Since the tests are performed with nominal size $\alpha = 0.05$, the rejection rate should be as close as possible to this value. It can be seen as the realized size of the tests.

For normal, skew-normal, and t-distributed error terms, all methods accurately estimate the indirect effect and the rejection rates of all bootstrap tests are close to the nominal size $\alpha = 0.05$. However, the OLS Sobel test is undersized for smaller effect sizes.

In the setting with outliers, ROBMED again yields the most accurate estimates of the indirect effect. Although there is some bias across the range of the effect size, it is far smaller than that of any other method. All other methods suffer from considerable bias, in particular the OLS-based methods. It is also noteworthy that the relative bias of all methods increases for large effect sizes of a, although that of ROBMED remains stable the longest. In addition, ROBMED is the only method with a rejection rate reasonably close to the nominal size $\alpha = 0.05$. While the rejection rate of the median bootstrap is not too far off the nominal size for smaller effect sizes of a, it increases slightly throughout the range of the effect size. All other tests have too large rejection rates, but interestingly the rejection rate of the winsorized bootstrap decreases with increasing effect size of a.[3]

Concluding discussion. The findings from the manuscript are robust across a range of effect sizes. In terms of estimating the indirect effect, ROBMED is the only method that is accurate across the four settings for error distributions and outliers. It is the most powerful test in the presence of outliers and heavy tails, and it does not lose much power under normal

[3] Winsorization moves outliers onto a certain tolerance ellipsoid of the estimated covariance matrix (here estimated via a Huber M-estimator; see Zu & Yuan, 2010). As the effect sizes of a and c become larger, the tolerance ellipsoids become more concentrated, meaning that the outliers are moved closer to the center for larger values of those effect sizes. That likely makes the influence of outliers decrease and the rejection rate move closer to the nominal size.

and skew-normal errors. Furthermore, it offers the best protection against falsely reporting mediation when the true indirect effect is 0.

Increasing sample size. We now discuss how an increasing sample size affects the results by comparing the results across Figures 1-10.

Simulations with mediation. From Figures 1-5, it is clear the OLS-based methods have insufficient power in the presence of outliers and heavy tails, even for large samples. As the sample size increases, the performance of the winsorized bootstrap, the median bootstrap, and the Box-Cox bootstrap becomes more similar to that of ROBMED under normal, skew-normal, and t-distributed errors. The most notable differences between methods are observed in the presence of outliers. Figure 11 therefore summarizes the results differently for this setting, with the sample size on the horizontal axis and separate columns for different effect sizes. Clearly, the Box-Cox transformation is not a suitable method to handle outliers. Hence the Box-Cox bootstrap shows a similar bias to the OLS-based methods, and it suffers from a severe loss of power for many effect sizes even when the sample size is large. ROBMED clearly performs the best for all sample sizes, followed by the winsorized bootstrap and the median bootstrap.

Simulations with no mediation. Figures 6-10 show that all bootstrap tests are fairly well calibrated (rejection rate close to the nominal size $\alpha = 0.05$), except for the setting with outliers. Figure 12 presents the results in the latter setting in a different manner, which reveals one of the most interesting results of this simulation design. As the sample size increases, across all effect sizes, ROBMED shows the smallest bias and it is the only method that is reasonably well calibrated. For all other methods, the rejection rate increases with increasing sample size. For $n = 1000$, the rejection rate of the other methods ranges from about 50% for the median bootstrap to almost 100% for the OLS bootstrap.

Concluding discussion. ROBMED is the only method with reliable results across all settings for error distributions and outliers, effect sizes, and sample sizes. In particular, ROBMED is the only method that protects against false mediation discoveries in the presence of outliers if the true indirect effect is 0, while overall being the most powerful across the different deviations from normality if the true indirect effect is nonzero.

Simulation design 2: Effect of outliers

We use the same basic simulation design from the manuscript, but we vary the outlier settings. The explanatory variable X and the error terms e_1 and e_2 are generated from independent standard normal distributions. We set $a = c = 0.4$, as well as $b = 0.4$ for a situation with mediation (true indirect effect $ab = 0.16$), and $b = 0$ for a situation where mediation does not exist (true indirect effect $ab = 0$). The parameters σ_1 and σ_2 are chosen such that M and Y have variance 1, and the sample sizes are $n = 100$ and $n = 250$. In addition to analyzing the clean data, we generate outliers in the following way. With probability $p = 0.01, 0.02, 0.03, 0.04$, observations are turned into outliers by setting $M_i^* = M_i/10 - d$ and $Y_i^* = Y_i/10 + d$, with outlier shift $d = 1, \dots, 6$. Note that the outlier setting from the manuscript is obtained by setting $p = 0.02$ and $d = 3$.

Simulations with mediation. Table 1 contains the results for the average estimates and standard errors of the indirect effect. For small values of the outlier shift d, the outliers overlap with the main data cloud, making it impossible for any method to distinguish the outliers from data points that follow the model. Hence all methods suffer from a bias in the estimate of the indirect effect for small values of d, with the bias increasing for higher outlier probabilities. As d becomes larger and the outliers become separable from the main data cloud, ROBMED is the only method for which the estimates recover from the bias. Its estimates move again closer to the true value and remain approximately unbiased. In addition to being the only method to recover from the bias, ROBMED also exhibits the smallest maximum bias (over the range of d). The estimates of the OLS-based methods, and to a lesser extent the estimates of the median bootstrap, continue to move away from the true value for an increasing outlier shift d. The winsorized bootstrap is able to stop the bias from increasing for reasonably small values of d, as winsorization cuts off the influence of outliers, but the bias remains constant and does not decrease again. For further illustration, Figure 13 (top) visualizes this behavior of the average estimates for outlier probability 2%.

Table 2 lists how often the methods reject the null hypothesis and the corresponding estimate of the indirect effect has the correct sign (our measure of realized power). The results from the estimation of the indirect effect clearly carry over. All tests except ROBMED continue to lose power for increasing outlier probability and increasing outlier shift d. Across all outlier probabilities, ROBMED only loses some power for small values of d when the outliers overlap with the main data cloud. As d increases further and the outliers become separable from the main part of the data, their power goes back to the same level that is observed on clean data ($d = 0$) and stays there for a broad range of d. Once its power stabilizes, ROBMED is the most powerful test. These results are further illustrated in Figure 13 (bottom) for outlier probability 2%.

Simulations with no mediation. The results for the average estimates and standard errors of the indirect effect are shown in Table 3. For increasing outlier probability and outlier shift d, the outliers push the estimates towards a negative estimated effect. Across all outlier probabilities, we observe the following behavior. ROBMED is again the only method for which the estimates recover from this bias and stay approximately unbiased once the outliers are separable from the main part of the data (large enough values of d). The winsorized bootstrap is able to contain the bias but is unable to decrease the bias again for larger values of the outlier shift d. The bias of the median bootstrap, and even more so the bias of the OLS-based methods, keeps increasing for increasing d. For outlier probability 2%, this behavior is also illustrated in Figure 14 (top).

The rejection rate of the tests (i.e., their realized size) is shown Table 4. The rejection rate for the OLS-based tests quickly rises with increasing outlier probability and increasing outlier shift d. The rejection rates of the Box-Cox bootstrap, the winsorized bootstrap, and the median bootstrap test increase (somewhat) more slowly. ROBMED exhibits a fairly stable rejection rate close to the nominal size $\alpha = 0.05$ except for a noticeable increase for smaller values of d where the outliers overlap with the main data cloud, which is more pronounced for higher outlier probabilities. A visualization of these results for outlier probability 2% can be found in Figure 14 (bottom).

Concluding discussion. ROBMED clearly outperforms the alternative methods in this simulation design. Across various outlier probabilities, it is the only method that can recover from bias and loss of power as the outliers become separable from the main data cloud. In addition, ROBMED does not lose much power to the OLS bootstrap when there are no outliers, and it is the only method to effectively protect against falsely detecting mediation when the true indirect effect is 0. Finally, the poor performance of the Box-Cox bootstrap demonstrates that nonlinear transformations are not a suitable treatment for outliers.

Simulation design 3: Effect of different error distributions

We again use the same basic simulation design from the manuscript, but this time we vary the error distributions for different levels of skewness and kurtosis. The explanatory variable X follows a standard normal distribution $N(0,1)$. We set $a = c = 0.4$, and $b = 0.4$ for a situation with mediation (true indirect effect $ab = 0.16$), as well as $b = 0$ for a situation where mediation does not exist (true indirect effect $ab = 0$). The parameters σ_1 and σ_2 are chosen such that M and Y have variance 1 in the case of normal error distributions, and the sample sizes are $n = 100$ and $n = 250$. We investigate the following three settings for the distributions of the error terms e_1 and e_2: (i) a centered log-normal distribution, (ii) skew-t distributions, and (iii) non-normal distributions generated via Fleishman's method (Fleishman, 1978).

Centered log-normal distribution. The error terms e_1 and e_2 follow a centered log-normal distribution. That is, the error terms are generated from a log-normal distribution $LogN(0,1)$ after which the expected value $e^{1/2}$ is subtracted. The skewness of this distribution is 6.185 and the excess kurtosis is 107.936, hence deviations from normality are quite severe.

Simulations with mediation. The top row of Figure 15 shows boxplots of the estimates of the indirect effect. Despite the strong deviation from normality, bias is very low for all methods. Keep in mind that the estimates of the Box-Cox bootstrap are not comparable to the true indirect effect due to the application of nonlinear transformations, which also demonstrates that nonlinear transformations are not always a suitable treatment for non-normality. ROBMED exhibits the smallest variance among all methods.

The bottom row of Figure 15 shows the rate of how often the methods reject the null hypothesis and the corresponding estimate of the indirect effect has the correct sign (our measure of realized power of the tests). For the smaller sample size ($n = 100$), ROBMED and the Box-Cox bootstrap are the only methods that have realized power of (almost) 100%. The winsorized bootstrap and the median bootstrap only have slightly lower power, but the OLS based tests perform poorly. For the larger sample size ($n = 250$), all robust tests have realized power of (almost) 100%, but the OLS-based tests still trail behind at about 80%.

Simulations with no mediation. Regarding the estimation of the indirect effect, the top row of Figure 16 shows that the bias is fairly low for all methods. ROBMED has again lower variance than the median bootstrap, the winsorized bootstrap, and the OLS-based methods. The bottom row of Figure 16 shows that the rejection rates of the Box-Cox bootstrap and the robust methods are close to the nominal size $\alpha = 0.05$ for the larger sample size ($n = 250$), with the rejection rate of the OLS bootstrap being somewhat too high at about 10%. For the smaller sample size ($n = 100$), all tests exhibit slightly elevated rejection rates, with the winsorized bootstrap having the highest rejection rate.

Concluding discussion. ROBMED and the Box-Cox bootstrap outperform the other tests in terms of power. Despite the severe deviations from normality, ROBMED estimates the indirect effect very accurately. On the other hand, the estimates of the Box-Cox bootstrap are not comparable with the true indirect effect due to the nonlinear transformations.

Skew-t distributions. The error terms e_1 and e_2 follow a skew-t distribution $ST(\xi, 1, \lambda, \upsilon)$[4] (e.g., Azzalini & Capitano, 2014). Note that a standard normal distribution is obtained for $\xi = 0$, $\lambda = 0$ and $\upsilon \to \infty$. Similarly, a t-distribution is obtained by setting $\xi = 0$ and $\lambda = 0$, and a skew-normal distribution is obtained for $\upsilon \to \infty$. However, interpretation of the parameters λ and υ is quite difficult, as their effect on skewness and kurtosis cannot be decoupled (Arellano-Valle & Azzalini, 2013). We set the parameter λ such that the skewness in case of $\upsilon \to \infty$ (skew-normal distribution) is $-0.995, -0.5, 0, 0.5, 0.995$,[5] and we set $\upsilon = \infty, 5, 2$. For $\upsilon = 2$, skewness and kurtosis are undefined, while $\upsilon = 5$ is the smallest integer value of υ for which both skewness and kurtosis are finite. The values of λ, υ, and the corresponding values of skewness and excess kurtosis are reported together with the simulation results in Tables 5–8. Finally, the location parameter ξ is chosen such that the rescaled error terms $\sigma_1 e_1$ and $\sigma_2 e_2$, respectively, have mean 0.

Simulations with mediation. The results for the estimates of the indirect effect in Table 5 show that the bias is close to 0 for all methods. We emphasize again that the estimates of the Box-Cox bootstrap are not comparable with the true indirect effect due to the nonlinear

[4] In the literature, the parameter λ of the skew-t distribution is usually denoted by α. We changed this notation here to avoid confusion with the significance level α of the mediation tests.

[5] The skewness of the skew-normal distribution is bounded by the interval $[-0.995, 0.995]$.

transformations. Table 6 reports how often the tests reject the null hypothesis and the corresponding estimate of the indirect effect has the correct sign (our measure of realized power). For most parameter settings, the differences in power between the bootstrap tests are small. For the smaller sample size ($n = 100$), the median bootstrap in general has (slightly) lower power than the other robust tests. Interestingly, all methods have lower power for symmetric error distributions with heavier tails ($\lambda = 0$, small υ) than for distributions with skewness and heavier tails ($\lambda \neq 0$, small υ). A likely explanation is that the variability of the skew-t distribution decreases as $|\lambda|$ increases, meaning that the uncertainty in the model is the highest for $\lambda = 0$.[6] The most pronounced differences among the methods are found for $\upsilon = 2$, where the tails are the heaviest: across all values of λ, ROBMED has slightly higher power than the other tests, while the OLS bootstrap suffers from a severe loss of power. For the larger sample size ($n = 250$), ROBMED, the winsorized bootstrap, the median bootstrap, and the Box-Cox bootstrap exhibit power of (close to) 100% across all parameter values, while the OLS bootstrap still suffers from a considerable loss of power for $\upsilon = 2$.

Simulations with no mediation. Tables 7 and 8 contain the results for the estimates of the indirect effect and the rejection rates of the tests, respectively. Differences between the methods are small. The bias is almost 0 for all methods. The rejection rates are close to the nominal size $\alpha = 0.05$ for most parameter settings, but all tests tend to slightly overreject the heavier the tails (i.e., the smaller υ).

Concluding discussion. For small samples, ROBMED outperforms the other methods, including the Box-Cox bootstrap. ROBMED has comparable power to other methods when deviations from normality are small, but (slightly) higher power when deviations are more severe.

[6] It is not possible to fix the variance of the skew-t distribution across the different parameter settings for λ and υ, as the variance of the skew-t distribution is infinite for $\upsilon = 2$.

Fleishman's method. We apply Fleishman's method (Fleishman, 1978) to generate the error terms e_1 and e_2 from non-normal distributions. Fleishman's method works as follows. For each variable, observations are first generated by a standard normal distribution, after which a polynomial transformation is applied. The parameters of this polynomial transformation are chosen such that the distribution of the transformed variable matches given values of skewness and kurtosis. However, such a polynomial transformation does not exist for all values of skewness and kurtosis, and Headrick & Kowalchuk (2007) note that not each polynomial transformation results in a so-called *valid* probability density function (pdf).[7] Accordingly, we only consider values of skewness and kurtosis that result in such a valid pdf. The investigated values of skewness and excess kurtosis are reported together with the simulation results in Tables 9–12.

Simulations with mediation. Table 9 contains the results for the estimates of the indirect effect, and Table 10 lists how often the tests reject the null hypothesis and the corresponding estimate of the indirect effect has the correct sign (our measure of realized power). As the deviations from normality are fairly small with Fleishman's method, the bias is close to 0 for all methods and power is (close to) 100%. For the smaller sample size ($n = 100$), only the median bootstrap suffers from loss of power for a small number of parameter settings, for instance with negative excess kurtosis.

Simulations with no mediation. The results for the estimates of the indirect effect and the rejection rates of the tests are shown in Tables 11 and 12, respectively. Bias is close to 0 for all methods, and rejection rates are similar for all tests. For the larger sample size ($n = 250$), all tests are well calibrated with rejection rates close to the nominal size $\alpha = 0.05$, but for the smaller sample size ($n = 100$) all tests have a tendency to slightly overreject.

[7] Among the conditions for a valid pdf is that percentage points and measures of central tendencies (such as mean, median and mode) can be computed (see Headrick & Kowalchuk, 2007).

Concluding discussion. Since deviations from normality are rather small, all tests perform very well. Only the median bootstrap suffers from loss of power for a small number of parameter settings.

Simulation design 4: Extensions of Zu & Yuan's simulation design

This simulation design consists of several extensions of the design of Zu & Yuan (2010). The explanatory variable X and the error terms e_1 and e_2 follow a standard normal distribution, and we set $\sigma_1 = \sigma_2 = 1$ such that the variance of the rescaled error terms remains 1. We set $a = c = 0.2$, but we vary the value of b to investigate different settings: $b = 0.2$ yields a setting with mediation (with a true indirect effect $ab = 0.04$), while $b = 0$ corresponds to a setting where mediation does not exist (true indirect effect $ab = 0$). The sample sizes are $n = 100$ and $n = 250$. In addition to analyzing the clean data, we replace a small percentage of observations with outliers by setting $M_i^* = M_i - d$ and $Y_i^* = Y_i + d$. The original design of Zu & Yuan (2010) is obtained by setting $n = 250$, $d = 6$, and by replacing 2, 4, 6, 8, 10 observations by outliers.

Effect of the amount of outliers. Similar to Zu & Yuan (2010), we set the outlier shift $d = 6$ and replace up to 4% of observations (in steps of 1 observation) with outliers.

Simulations with mediation. In the top row of Figure 17, the average estimates of the indirect effect are shown for an increasing percentage of outliers. Clearly, the OLS-based methods show a large bias for the indirect effect in the presence of outliers, with this bias continuously increasing for an increasing percentage of outliers. The median bootstrap and the winsorized bootstrap are affected to a lesser extent, but also their bias keeps increasing as the percentage of outliers increases. However, ROBMED remains stable and accurate in estimating the indirect effect.

The bottom row of Figure 17 displays the rate of how often the methods reject the null hypothesis and the corresponding estimate of the indirect effect has the correct sign (our measure of realized power). For the OLS-based tests, the results from the estimation of the indirect effect clearly carry over. Their realized power quickly drops and reaches 0 when there are about 1% of outliers. The Box-Cox bootstrap, the median bootstrap, and the winsorized bootstrap also continuously lose power, and eventually their realized power drops to (almost) 0. ROBMED is the only test that remains stable. It is the most powerful test when there are about 1% of outliers or more.

Simulations with no mediation. In the top row of Figure 18, we observe that the outliers push the estimates of the OLS-based methods towards a negative estimated effect. A similar effect, although to a lesser extent, is visible for the estimates of the median bootstrap and the winsorized bootstrap. ROBMED, on the other hand, remains stable and close to the true value $ab = 0$.

The bottom row of Figure 18 presents the rejection rate of the tests (i.e., their realized size). As expected, the rejection rate for the OLS-based tests quickly rises, but interestingly it starts to fall again for higher percentages of outliers. This is likely because of the estimated confidence intervals being even more affected by the outliers than the point estimates, yielding very large confidence intervals for higher percentages of outliers. The rejection rates of the Box-Cox bootstrap, the winsorized bootstrap, and the median bootstrap increase (somewhat) more slowly. ROBMED is the only test unaffected by the outliers and its rejection rate remains close to the nominal size $\alpha = 0.05$.

Concluding discussion. ROBMED clearly outperforms the alternative methods in this simulation design. It remains accurate in estimating the indirect effect and powerful for assessing its significance. In addition, ROBMED does not lose much power to the OLS bootstrap when there are no outliers. ROBMED's competitors show a loss of power in the presence of outliers, and can be driven to falsely report mediation when the true indirect effect is 0.

Effect of the distance of outliers from the main bulk of the data. We fix the percentage of outliers to 2% and we vary the outlier shift $d = 0, \dots, 15$.

Simulations with mediation. The top row of Figure 19 displays the average estimates of the indirect effect for varying values of the outlier shift d. For small values of this shift, where outliers overlap with the main bulk of the data, the bias of all methods increases. As the outlier shift d increases further and the outliers become separable from the main bulk of the data, ROBMED is the only methods for which the bias goes back towards zero. For the OLS bootstrap, and to a lesser extent for the median bootstrap, the bias keeps increasing. For the winsorized bootstrap, the bias stabilizes but does not decrease again.

The bottom row of Figure 19 shows how often the methods reject the null hypothesis and the corresponding estimate of the indirect effect has the correct sign (our measure of realized power). As in the previous simulation design, the results from the estimation of the indirect effect carry over. All tests except ROBMED continue to lose power for an increasing outlier shift d. ROBMED only loses some power for small values of d when the outliers overlap with the main data cloud. As d increases further and the outliers become separable from the main part of the data, their power goes back to the same level that is observed on clean data ($d = 0$) and stays there for a broad range of d. Once its power stabilizes, ROBMED is the most powerful test.

Simulations with no mediation. From the top row of Figure 20, it is clear that the outliers push the estimates towards a negative estimated effect. Otherwise, the results are pretty similar to the case with mediation. ROBMED is again the only method for which the estimates recover from this bias and stay approximately unbiased once the outliers are separable from the main part of the data. The winsorized bootstrap is able to contain the bias but is unable to decrease the bias again for larger values of the outlier shift d. The bias of the median bootstrap, and even more so the bias of the OLS-based methods, keeps increasing for increasing d.

The rejection rate of the tests (i.e., their realized size) is shown in the bottom row of Figure 20. As in the previous simulation design, the rejection rate for the OLS-based tests quickly rises, but starts to fall again for larger values of d. An explanation could again be that the estimated confidence intervals are even more affected by large outliers than the point estimates. The rejection rates of the Box-Cox bootstrap, the winsorized bootstrap, and the median bootstrap test increase (somewhat) more slowly. While the rejection rate of the winsorized bootstrap test levels off for reasonably small d, and to a lesser extent that of the Box-Cox bootstrap as well, the rejection rate of the median bootstrap keeps increasing. ROBMED exhibits a fairly stable rejection rate close to the nominal size $\alpha = 0.05$ except for a small bump for smaller values of d where the outliers overlap with the main data cloud.

Concluding discussion. ROBMED shows the best overall performance in this simulation design. It is the only method that can recover from bias and loss of power as the outliers become separable from the main data cloud. Furthermore, it is the only method to effectively protect against falsely detecting mediation when the true indirect effect is 0.

Simulation design 5: Yuan & MacKinnon's simulation design

The last simulation design is taken from Yuan & MacKinnon (2014). First, the explanatory variable X is generated from a standard normal distribution. We investigate several settings where mediation exists: we set $a = b = c = 0.14$ for small effect sizes (yielding a true indirect effect $ab = 0.0196$), $a = b = c = 0.39$ for medium effect sizes (yielding a true indirect effect $ab = 0.1521$), and $a = b = c = 0.59$ for larger effect sizes (yielding a true indirect effect $ab = 0.3481$). To study situations where mediation does not exist, we keep the same values for b and c, but set $a = 0$ for a true indirect effect $ab = 0$.

The sample sizes are $n = 50, 100, 200, 500$. Moreover, we set $\sigma_1 = \sigma_2 = 1$ and consider the following distributions of the error terms e_1 and e_2:

1. A standard normal distribution $N(0,1)$.
2. A t distribution with 2 degrees of freedom as an example of a distribution with heavy tails.
3. A contaminated normal distribution $0.9 \cdot N(0,1) + 0.1 \cdot N(0, 10^2)$, i.e., with 10% probability, error terms are generated from a normal distribution with a much larger variance.

Even though results for all effect sizes are shown in Figures 21-28, we discuss only the results for the medium effect sizes, as the results for the smaller and larger effect sizes are qualitatively similar. The main difference is that, as expected, all methods have lower power for the smaller effect sizes and higher power for the larger effect sizes.

Simulations with mediation. Figures 21–24 show the simulation results for sample size $n = 50, 100, 200$ and 500, respectively, and the following discussion focuses on the setting with $a = b = c = 0.39$ and true indirect effect $ab = 0.1521$. The top row of the figures contains box plots of the estimates of the indirect effect, while the bottom row displays rate of how often the methods reject the null hypothesis and the corresponding estimate of ab has the correct sign (i.e., the realized power of the tests). The columns of the figures correspond to the three considered distributions of the error terms.

For standard normal error terms, all methods estimate the indirect effect accurately. As expected, the power is the highest for the OLS bootstrap and the winsorized bootstrap, followed by ROBMED and the median bootstrap. The power of all bootstrap tests increases with increasing sample size, and all tests already reach a 100% rejection rate for $n = 200$ (or are at least very close).

For heavy tails in the errors (following a t distribution with 2 degrees of freedom), the Box-Cox bootstrap estimates and the OLS-based estimates of the indirect effect show larger variability than the other methods, and the OLS-based tests have the lowest power. The other methods are very similar in terms of estimating the indirect effect and remain accurate. ROBMED thereby has the highest power, followed by the winsorized bootstrap, the Box-Cox bootstrap, and the median bootstrap.

When the error terms are generated by a contaminated normal distribution, the results are very similar to the setting with heavy tails. The OLS-based estimates and the Box-Cox bootstrap estimates again show larger variability, and the OLS-based tests have the lowest power. As before, ROBMED has the highest power, followed by the winsorized bootstrap, the Box-Cox bootstrap and the median bootstrap.

Simulations with no mediation. Figures 25–28 show the simulation results for sample size $n = 50, 100, 200$ and 500, respectively, and the following discussion focuses on the setting with $a = 0$, $b = c = 0.39$ and true indirect effect $ab = 0$. The top row of the figures again shows box plots of the estimates of the indirect effect, while the bottom row displays the rejection rate (i.e., the realized size of the tests). For the most part, all methods perform fairly similarly. The most interesting result is that the OLS-based estimates and the Box-Cox bootstrap estimates again show larger variability for errors with heavy tails and errors from a contaminated normal distribution. In addition, the realized size of the OLS bootstrap test is slightly elevated under those two error distributions.

Concluding discussion. There are fewer differences between the methods than in other simulation designs. ROBMED outperforms its competitors in the setting with heavy tails in the errors, as well as in the setting where the error terms are generated by a contaminated normal distribution.

Online Appendix 3: Literature Review

We conducted a review of empirical articles that tested mediation, published in Academy of Management Journal, Strategic Management Journal, Journal of Applied Psychology, Organization Science, and Administrative Science Quarterly in 2019. Table 13 reports whether these articles (i) use OLS, (ii) report outliers, (iii) check model assumptions (e.g., normality), (iv) use bootstrapping, and (v) use the PROCESS macro (Hayes, 2018) (or its earlier versions).

References

Arellano-Valle, R. B., & Azzalini, A. (2013). The Centred Parameterization and Related Quantiles of the Skew-t distribution. *Journal of Multivariate Analysis, 113*, 73–90.

Azzalini, A., & Capitano, A. (2014). *Azzalini, A., & Capitanio, A. (2014). The Skew-Normal and Related Families.* Cambridge, UK: Cambridge University Press.

Becker, T. E., Robertson, M. M., & Vandenberg, R. J. (2019). Nonlinear Transformations in Organizational Research: Possible Problems and Potential Solutions. *Organizational Research Methods, 22*(4), 831–866.

Bishop, J. W., & Scott, K. W. (2000). An Examination of Organizational and Team Commitment in a Self-Directed Team Environment. *Journal of Applied Psychology, 85*(3), 439–450.

Bollen, K. A., & Stine, R. (1990). Direct and Indirect Effects: Classical and Bootstrap Estimates of Variability. *Sociological Methodology, 20*, 115–140.

Box, G. E., & Cox, D. R. (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society, Series B, 26*(2), 211–252.

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap Methods and their Application.* Cambridge, UK: Cambridge University Press.

Fleishman, A. I. (1978). A Method for Simulating Non-Normal Distributions. *Psychometrika, 43*(4), 521–532.

Hawkins, D. M., & Weisberg, S. (2017). Combining the Box-Cox Power and Generalized Log Transformations to Accomodate Nonpositive Responses In Linear and Mixed-Effects Linear Models. *South African Statistical Journal, 51*(2), 317–328.

Hayes, A. F. (2018). *Introduction to Mediation, Moderation, and Conditional Process Analysis* (2nd ed.). New York: The Guilford Press.

Headrick, T. C., & Kowalchuk, R. K. (2007). The Power Method Transformation: Its Probability Density Function, and Its Further Use for Fitting Data. *Journal of Statistical Computation and Simulation, 77*(3), 229–249.

Jehn, K. A. (1994). Enhancing Effectiveness: An Investigation of Advantages and Disadvantages of Value-Based Intragroup Conflict. *International Journal of Conflict Management, 5*(3), 223–238.

Jehn, K. A. (1995). A Multi-Method Examination of the Benefits and Detriments of Intra-Group Conflict. *Administrative Science Quarterly, 40*(2), 256–285.

Jehn, K. A., Northcraft, G. B., & Neale, M. A. (1999). Why Differences Make a Difference: A Field Study of Diversity, Conflict and Performance in Workgroups. *Administrative Science Quarterly, 44*(4), 741–763.

Lindeman, M., & Verkasalo, M. (2005). Measuring Values With the Short Schwartz's Value Survey. *Journal of Personality Assessment, 85*(2), 170–178.

MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence Limits for the Indirect Effect: Distribution of the Product and Resampling Methods. *Multivariate Behavioral Research, 39*(1), 99–128.

Mowday, R. T., Steers, R. M., & Porter, L. W. (1979). The Measurement of Organizational Commitment. *Journal of Vocational Behavior, 14*(2), 224–247.

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS Procedures for Estimating Indirect Effects in Simple Mediation Models. *Bahavior Research Methods, Instruments, & Computers, 36*(4), 717–731.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and Resampling Strategies for Assessing and Comparing Indirect Effects in Multiple Mediator Models. *Behavior Research Methods, 40*(3), 879–891.

Rokeach, M. (1973). *The Nature of Human Values.* New York, NY: The Free Press.

Salibián-Barrera, M., & Zamar, R. H. (2002). Bootstrapping Robust Estimates of Regression. *The Annals of Statistics, 30*(2), 556–582.

Schwartz, S. H. (1992). Universals in the Content and Structure of Values: Theoretical Advances and Empirical Tests in 20 Countries. *Advances in Experimental Social Psychology, 25*, 1–65.

Shrout, P. E., & Bolger, N. (2002). Mediation in Experimental and Nonexperimental Studies: New Procedures and Recommendations. *Psychological Methods, 7*(4), 422–445.

Sobel, M. E. (1982). Asymptotic Confidence Intervals for Indirect Effects in Structural Equation Models. *Sociological Methodology, 13*, 290–312.

Yohai, V. J. (1987). High Breakdown-Point and High Efficiency Robust Estimates for Regression. *The Annals of Statistics, 15*(2), 642–656.

Yuan, Y., & MacKinnon, D. P. (2014). Robust Mediation Analysis Based on Median Regression. *Psychological Methods, 19*(1), 1–20.

Zu, J., & Yuan, K.-H. (2010). Local Influence and Robust Procedures for Mediation Analysis. *Multivariate Behavioral Research, 45*(1), 1–44.

Tables

Table 1. Bias and standard deviation (in parenthesis) for simulation design 2 with outliers in the setting with mediation ($a = b = 0.4$).

n	Probability of outliers	Outlier shift d	OLS bootstrap	OLS Sobel	Box-Cox bootstrap	Winsorized bootstrap	Median bootstrap	ROBMED
100	0%	0	0.002 (0.052)	0.002 (0.052)	0.006 (0.066)	0.002 (0.052)	0.002 (0.059)	0.002 (0.056)
	1%	1	-0.006 (0.051)	-0.006 (0.051)	-0.002 (0.064)	-0.006 (0.052)	-0.004 (0.058)	-0.005 (0.056)
	1%	2	-0.024 (0.053)	-0.024 (0.053)	-0.021 (0.063)	-0.012 (0.053)	-0.009 (0.058)	0.001 (0.058)
	1%	3	-0.048 (0.063)	-0.051 (0.065)	-0.044 (0.066)	-0.012 (0.053)	-0.014 (0.059)	0.003 (0.057)
	1%	4	-0.073 (0.079)	-0.080 (0.083)	-0.050 (0.071)	-0.013 (0.053)	-0.019 (0.060)	0.002 (0.056)
	1%	5	-0.097 (0.097)	-0.110 (0.104)	-0.044 (0.072)	-0.013 (0.053)	-0.025 (0.062)	0.002 (0.056)
	1%	6	-0.118 (0.114)	-0.137 (0.125)	-0.039 (0.072)	-0.013 (0.053)	-0.031 (0.064)	0.002 (0.056)
	2%	1	-0.013 (0.050)	-0.013 (0.050)	-0.010 (0.062)	-0.014 (0.051)	-0.009 (0.058)	-0.012 (0.055)
	2%	2	-0.047 (0.053)	-0.048 (0.054)	-0.044 (0.059)	-0.027 (0.054)	-0.019 (0.058)	0.000 (0.058)
	2%	3	-0.089 (0.066)	-0.093 (0.068)	-0.076 (0.059)	-0.028 (0.055)	-0.030 (0.059)	0.004 (0.057)
	2%	4	-0.131 (0.083)	-0.141 (0.086)	-0.082 (0.065)	-0.029 (0.055)	-0.041 (0.062)	0.002 (0.056)
	2%	5	-0.169 (0.101)	-0.186 (0.104)	-0.075 (0.069)	-0.029 (0.055)	-0.053 (0.067)	0.002 (0.056)
	2%	6	-0.202 (0.118)	-0.226 (0.121)	-0.070 (0.072)	-0.029 (0.055)	-0.066 (0.074)	0.002 (0.056)
	3%	1	-0.020 (0.049)	-0.020 (0.049)	-0.019 (0.060)	-0.022 (0.050)	-0.015 (0.057)	-0.019 (0.054)
	3%	2	-0.068 (0.053)	-0.070 (0.054)	-0.065 (0.056)	-0.045 (0.057)	-0.030 (0.058)	-0.002 (0.060)
	3%	3	-0.125 (0.068)	-0.130 (0.068)	-0.102 (0.054)	-0.048 (0.059)	-0.047 (0.061)	0.005 (0.058)
	3%	4	-0.179 (0.085)	-0.190 (0.086)	-0.108 (0.062)	-0.049 (0.061)	-0.065 (0.067)	0.002 (0.056)
	3%	5	-0.226 (0.102)	-0.242 (0.101)	-0.104 (0.070)	-0.050 (0.061)	-0.085 (0.077)	0.002 (0.056)
	3%	6	-0.264 (0.118)	-0.286 (0.116)	-0.102 (0.076)	-0.051 (0.062)	-0.107 (0.089)	0.002 (0.056)
	4%	1	-0.027 (0.048)	-0.027 (0.048)	-0.025 (0.058)	-0.029 (0.049)	-0.020 (0.056)	-0.025 (0.053)
	4%	2	-0.087 (0.053)	-0.088 (0.053)	-0.083 (0.054)	-0.064 (0.059)	-0.041 (0.058)	-0.005 (0.063)
	4%	3	-0.155 (0.068)	-0.160 (0.068)	-0.122 (0.051)	-0.070 (0.066)	-0.065 (0.063)	0.007 (0.058)
	4%	4	-0.217 (0.085)	-0.227 (0.084)	-0.130 (0.062)	-0.073 (0.070)	-0.091 (0.074)	0.002 (0.056)
	4%	5	-0.268 (0.103)	-0.283 (0.100)	-0.131 (0.074)	-0.075 (0.073)	-0.120 (0.090)	0.002 (0.056)
	4%	6	-0.310 (0.120)	-0.328 (0.116)	-0.133 (0.082)	-0.076 (0.075)	-0.150 (0.105)	0.002 (0.056)
250	0%	0	0.002 (0.032)	0.002 (0.032)	0.001 (0.045)	0.002 (0.032)	0.002 (0.037)	0.001 (0.035)
	1%	1	-0.006 (0.032)	-0.006 (0.032)	-0.006 (0.043)	-0.006 (0.032)	-0.003 (0.037)	-0.005 (0.035)
	1%	2	-0.024 (0.034)	-0.024 (0.034)	-0.024 (0.043)	-0.011 (0.033)	-0.008 (0.037)	0.002 (0.036)
	1%	3	-0.050 (0.041)	-0.052 (0.042)	-0.049 (0.045)	-0.012 (0.033)	-0.012 (0.037)	0.002 (0.035)
	1%	4	-0.081 (0.053)	-0.085 (0.054)	-0.055 (0.050)	-0.012 (0.033)	-0.017 (0.038)	0.001 (0.035)
	1%	5	-0.113 (0.066)	-0.120 (0.068)	-0.043 (0.052)	-0.012 (0.033)	-0.022 (0.039)	0.001 (0.035)
	1%	6	-0.144 (0.079)	-0.154 (0.081)	-0.035 (0.052)	-0.012 (0.033)	-0.027 (0.040)	0.001 (0.035)
	2%	1	-0.013 (0.031)	-0.013 (0.031)	-0.013 (0.042)	-0.014 (0.032)	-0.008 (0.037)	-0.012 (0.034)
	2%	2	-0.048 (0.034)	-0.049 (0.034)	-0.048 (0.040)	-0.027 (0.034)	-0.018 (0.037)	0.002 (0.037)
	2%	3	-0.095 (0.043)	-0.097 (0.043)	-0.081 (0.039)	-0.028 (0.034)	-0.027 (0.037)	0.004 (0.036)
	2%	4	-0.145 (0.054)	-0.150 (0.054)	-0.080 (0.043)	-0.028 (0.034)	-0.037 (0.038)	0.001 (0.035)
	2%	5	-0.193 (0.064)	-0.201 (0.064)	-0.069 (0.047)	-0.028 (0.034)	-0.048 (0.040)	0.001 (0.035)
	2%	6	-0.236 (0.074)	-0.247 (0.073)	-0.063 (0.049)	-0.028 (0.034)	-0.058 (0.044)	0.001 (0.035)
	3%	1	-0.020 (0.031)	-0.020 (0.031)	-0.019 (0.040)	-0.021 (0.031)	-0.013 (0.037)	-0.018 (0.034)
	3%	2	-0.068 (0.033)	-0.069 (0.033)	-0.067 (0.037)	-0.043 (0.035)	-0.027 (0.037)	0.001 (0.038)
	3%	3	-0.130 (0.041)	-0.132 (0.042)	-0.104 (0.034)	-0.044 (0.036)	-0.042 (0.038)	0.005 (0.036)
	3%	4	-0.191 (0.051)	-0.196 (0.051)	-0.103 (0.041)	-0.045 (0.036)	-0.057 (0.040)	0.001 (0.035)
	3%	5	-0.246 (0.060)	-0.253 (0.059)	-0.096 (0.046)	-0.046 (0.036)	-0.073 (0.044)	0.001 (0.035)
	3%	6	-0.293 (0.068)	-0.302 (0.066)	-0.094 (0.050)	-0.046 (0.036)	-0.091 (0.051)	0.001 (0.035)
	4%	1	-0.026 (0.030)	-0.026 (0.030)	-0.026 (0.039)	-0.028 (0.030)	-0.018 (0.036)	-0.024 (0.033)
	4%	2	-0.087 (0.032)	-0.088 (0.032)	-0.085 (0.035)	-0.061 (0.037)	-0.037 (0.036)	-0.001 (0.039)
	4%	3	-0.160 (0.041)	-0.162 (0.041)	-0.124 (0.032)	-0.064 (0.039)	-0.057 (0.038)	0.006 (0.037)
	4%	4	-0.228 (0.050)	-0.232 (0.049)	-0.127 (0.041)	-0.066 (0.040)	-0.079 (0.043)	0.002 (0.035)
	4%	5	-0.286 (0.059)	-0.292 (0.058)	-0.126 (0.049)	-0.067 (0.041)	-0.103 (0.052)	0.001 (0.035)
	4%	6	-0.333 (0.069)	-0.339 (0.067)	-0.127 (0.054)	-0.067 (0.041)	-0.130 (0.063)	0.001 (0.035)

Table 2. Rate of rejection with correct sign of the indirect effect (realized power; the higher the better) for simulation design 2 with outliers in the setting with mediation ($a = b = 0.4$).

n	Probability of outliers	Outlier shift d	OLS bootstrap	OLS Sobel	Box-Cox bootstrap	Winsorized bootstrap	Median bootstrap	ROBMED
100	0%	0	0.984	0.957	0.979	0.985	0.842	0.958
	1%	1	0.976	0.953	0.973	0.978	0.829	0.942
	1%	2	0.875	0.868	0.888	0.965	0.797	0.935
	1%	3	0.456	0.670	0.633	0.963	0.748	0.952
	1%	4	0.363	0.469	0.477	0.960	0.687	0.954
	1%	5	0.363	0.364	0.414	0.961	0.627	0.954
	1%	6	0.363	0.357	0.395	0.961	0.563	0.954
	2%	1	0.969	0.937	0.966	0.972	0.808	0.932
	2%	2	0.701	0.715	0.739	0.901	0.727	0.903
	2%	3	0.219	0.389	0.367	0.892	0.633	0.945
	2%	4	0.122	0.222	0.232	0.886	0.536	0.948
	2%	5	0.122	0.132	0.169	0.886	0.435	0.948
	2%	6	0.122	0.121	0.154	0.884	0.349	0.948
	3%	1	0.964	0.919	0.955	0.966	0.789	0.911
	3%	2	0.503	0.540	0.564	0.803	0.660	0.864
	3%	3	0.094	0.209	0.209	0.771	0.498	0.944
	3%	4	0.044	0.095	0.107	0.760	0.374	0.947
	3%	5	0.044	0.048	0.069	0.750	0.268	0.948
	3%	6	0.044	0.043	0.062	0.748	0.199	0.948
	4%	1	0.960	0.903	0.946	0.960	0.748	0.901
	4%	2	0.367	0.388	0.434	0.669	0.566	0.814
	4%	3	0.035	0.102	0.112	0.624	0.394	0.944
	4%	4	0.015	0.037	0.047	0.608	0.264	0.946
	4%	5	0.015	0.016	0.026	0.602	0.168	0.947
	4%	6	0.015	0.014	0.024	0.598	0.103	0.947
250	0%	0	1	1	1	1	1	1
	1%	1	1	1	1	1	1	1
	1%	2	1	1	1	1	0.999	1
	1%	3	0.767	0.940	0.954	1	0.997	1
	1%	4	0.260	0.686	0.849	1	0.990	1
	1%	5	0.090	0.437	0.757	1	0.971	1
	1%	6	0.089	0.272	0.688	1	0.927	1
	2%	1	1	1	1	1	0.999	1
	2%	2	0.977	0.987	0.983	0.998	0.996	1
	2%	3	0.325	0.650	0.732	0.997	0.978	1
	2%	4	0.035	0.212	0.497	0.995	0.905	1
	2%	5	0.009	0.085	0.376	0.995	0.787	1
	2%	6	0.009	0.035	0.302	0.995	0.650	1
	3%	1	1	1	1	1	0.998	1
	3%	2	0.876	0.929	0.920	0.983	0.983	0.998
	3%	3	0.089	0.310	0.452	0.973	0.906	1
	3%	4	0.003	0.046	0.208	0.966	0.743	1
	3%	5	0	0.012	0.129	0.962	0.544	1
	3%	6	0	0.004	0.088	0.959	0.349	1
	4%	1	1	1	1	1	0.998	1
	4%	2	0.721	0.797	0.782	0.919	0.951	0.987
	4%	3	0.020	0.101	0.200	0.881	0.791	1
	4%	4	0	0.009	0.057	0.863	0.529	1
	4%	5	0	0.001	0.034	0.857	0.292	1
	4%	6	0	0	0.020	0.851	0.150	1

Table 3. Bias and standard deviation (in parenthesis) for simulation design 2 with outliers in the setting with no mediation ($a = 0.4$, $b = 0$).

n	Probability of outliers	Outlier shift d	OLS bootstrap	OLS Sobel	Box-Cox bootstrap	Winsorized bootstrap	Median bootstrap	ROBMED
100	0%	0	0.002 (0.042)	0.002 (0.042)	0.003 (0.044)	0.002 (0.042)	0.001 (0.049)	0.001 (0.047)
	1%	1	-0.002 (0.042)	-0.002 (0.042)	-0.002 (0.044)	-0.003 (0.042)	-0.005 (0.049)	-0.004 (0.046)
	1%	2	-0.015 (0.044)	-0.016 (0.044)	-0.014 (0.046)	-0.014 (0.045)	-0.012 (0.052)	-0.011 (0.051)
	1%	3	-0.032 (0.052)	-0.034 (0.053)	-0.021 (0.047)	-0.016 (0.047)	-0.018 (0.054)	-0.006 (0.055)
	1%	4	-0.050 (0.065)	-0.055 (0.067)	-0.027 (0.056)	-0.016 (0.048)	-0.025 (0.058)	0.000 (0.051)
	1%	5	-0.067 (0.079)	-0.076 (0.084)	-0.036 (0.066)	-0.017 (0.048)	-0.031 (0.062)	0.001 (0.049)
	1%	6	-0.083 (0.093)	-0.096 (0.100)	-0.042 (0.073)	-0.017 (0.048)	-0.038 (0.067)	0.001 (0.048)
	2%	1	-0.006 (0.041)	-0.006 (0.041)	-0.006 (0.044)	-0.007 (0.042)	-0.010 (0.049)	-0.009 (0.046)
	2%	2	-0.030 (0.046)	-0.031 (0.046)	-0.028 (0.047)	-0.028 (0.048)	-0.024 (0.054)	-0.024 (0.055)
	2%	3	-0.060 (0.057)	-0.063 (0.058)	-0.039 (0.050)	-0.034 (0.052)	-0.037 (0.059)	-0.017 (0.064)
	2%	4	-0.090 (0.072)	-0.097 (0.074)	-0.052 (0.066)	-0.035 (0.053)	-0.050 (0.065)	-0.003 (0.058)
	2%	5	-0.117 (0.088)	-0.129 (0.090)	-0.068 (0.080)	-0.036 (0.054)	-0.064 (0.073)	0.000 (0.052)
	2%	6	-0.140 (0.104)	-0.157 (0.105)	-0.080 (0.089)	-0.037 (0.054)	-0.077 (0.081)	0.001 (0.049)
	3%	1	-0.010 (0.041)	-0.010 (0.041)	-0.011 (0.043)	-0.011 (0.041)	-0.016 (0.050)	-0.015 (0.046)
	3%	2	-0.045 (0.047)	-0.046 (0.047)	-0.041 (0.049)	-0.043 (0.049)	-0.037 (0.056)	-0.039 (0.059)
	3%	3	-0.085 (0.061)	-0.089 (0.061)	-0.055 (0.054)	-0.055 (0.059)	-0.057 (0.065)	-0.032 (0.079)
	3%	4	-0.124 (0.077)	-0.132 (0.077)	-0.075 (0.078)	-0.058 (0.061)	-0.078 (0.074)	-0.010 (0.072)
	3%	5	-0.157 (0.094)	-0.169 (0.093)	-0.100 (0.095)	-0.059 (0.062)	-0.098 (0.085)	-0.003 (0.058)
	3%	6	-0.185 (0.110)	-0.201 (0.109)	-0.117 (0.106)	-0.060 (0.063)	-0.118 (0.094)	-0.001 (0.055)
	4%	1	-0.014 (0.040)	-0.014 (0.040)	-0.015 (0.043)	-0.015 (0.041)	-0.021 (0.050)	-0.020 (0.045)
	4%	2	-0.057 (0.048)	-0.058 (0.047)	-0.053 (0.050)	-0.057 (0.050)	-0.050 (0.059)	-0.054 (0.062)
	4%	3	-0.106 (0.064)	-0.110 (0.063)	-0.067 (0.056)	-0.075 (0.065)	-0.077 (0.070)	-0.050 (0.092)
	4%	4	-0.151 (0.082)	-0.158 (0.081)	-0.094 (0.087)	-0.081 (0.071)	-0.105 (0.083)	-0.020 (0.088)
	4%	5	-0.188 (0.101)	-0.198 (0.099)	-0.126 (0.106)	-0.084 (0.073)	-0.132 (0.095)	-0.006 (0.070)
	4%	6	-0.218 (0.120)	-0.230 (0.118)	-0.148 (0.118)	-0.085 (0.075)	-0.157 (0.104)	-0.004 (0.064)
250	0%	0	0.002 (0.026)	0.002 (0.026)	0.002 (0.027)	0.001 (0.027)	0.001 (0.031)	0.001 (0.029)
	1%	1	-0.003 (0.026)	-0.002 (0.026)	-0.003 (0.027)	-0.003 (0.026)	-0.004 (0.031)	-0.004 (0.029)
	1%	2	-0.016 (0.028)	-0.016 (0.028)	-0.015 (0.029)	-0.014 (0.029)	-0.010 (0.033)	-0.011 (0.032)
	1%	3	-0.035 (0.034)	-0.036 (0.034)	-0.024 (0.029)	-0.015 (0.029)	-0.016 (0.034)	-0.004 (0.033)
	1%	4	-0.057 (0.042)	-0.059 (0.043)	-0.033 (0.037)	-0.016 (0.030)	-0.022 (0.036)	0.001 (0.030)
	1%	5	-0.079 (0.052)	-0.084 (0.053)	-0.044 (0.044)	-0.016 (0.030)	-0.028 (0.038)	0.001 (0.029)
	1%	6	-0.102 (0.062)	-0.109 (0.063)	-0.052 (0.049)	-0.017 (0.030)	-0.034 (0.040)	0.001 (0.029)
	2%	1	-0.007 (0.026)	-0.007 (0.026)	-0.008 (0.027)	-0.008 (0.026)	-0.010 (0.031)	-0.010 (0.029)
	2%	2	-0.032 (0.028)	-0.032 (0.028)	-0.030 (0.029)	-0.030 (0.030)	-0.022 (0.034)	-0.025 (0.034)
	2%	3	-0.066 (0.036)	-0.067 (0.036)	-0.044 (0.031)	-0.034 (0.032)	-0.035 (0.036)	-0.012 (0.040)
	2%	4	-0.102 (0.045)	-0.105 (0.046)	-0.064 (0.046)	-0.036 (0.033)	-0.047 (0.040)	0.000 (0.031)
	2%	5	-0.136 (0.055)	-0.142 (0.055)	-0.085 (0.055)	-0.037 (0.033)	-0.060 (0.045)	0.001 (0.029)
	2%	6	-0.167 (0.064)	-0.175 (0.064)	-0.099 (0.061)	-0.037 (0.033)	-0.073 (0.050)	0.001 (0.029)
	3%	1	-0.010 (0.026)	-0.010 (0.026)	-0.012 (0.027)	-0.011 (0.026)	-0.015 (0.032)	-0.014 (0.029)
	3%	2	-0.045 (0.029)	-0.046 (0.029)	-0.042 (0.030)	-0.044 (0.030)	-0.034 (0.035)	-0.039 (0.037)
	3%	3	-0.089 (0.037)	-0.091 (0.037)	-0.058 (0.034)	-0.053 (0.035)	-0.052 (0.040)	-0.024 (0.050)
	3%	4	-0.133 (0.047)	-0.137 (0.046)	-0.089 (0.053)	-0.056 (0.037)	-0.071 (0.046)	-0.002 (0.037)
	3%	5	-0.173 (0.056)	-0.178 (0.056)	-0.118 (0.063)	-0.057 (0.037)	-0.091 (0.053)	0.001 (0.029)
	3%	6	-0.207 (0.066)	-0.213 (0.065)	-0.137 (0.070)	-0.058 (0.038)	-0.111 (0.061)	0.001 (0.029)
	4%	1	-0.014 (0.025)	-0.014 (0.025)	-0.016 (0.027)	-0.015 (0.025)	-0.020 (0.031)	-0.019 (0.028)
	4%	2	-0.058 (0.029)	-0.058 (0.029)	-0.054 (0.031)	-0.058 (0.031)	-0.046 (0.036)	-0.054 (0.038)
	4%	3	-0.110 (0.038)	-0.111 (0.038)	-0.071 (0.036)	-0.074 (0.040)	-0.071 (0.043)	-0.042 (0.062)
	4%	4	-0.159 (0.049)	-0.162 (0.049)	-0.111 (0.060)	-0.078 (0.042)	-0.097 (0.052)	-0.005 (0.047)
	4%	5	-0.200 (0.060)	-0.204 (0.060)	-0.148 (0.071)	-0.080 (0.043)	-0.125 (0.062)	0.001 (0.029)
	4%	6	-0.234 (0.072)	-0.239 (0.071)	-0.171 (0.078)	-0.082 (0.043)	-0.153 (0.071)	0.001 (0.029)

Table 4. Rejection rate (realized size of the tests; the closer to $\alpha = 0.05$ the better) for simulation design 2 with outliers in the setting with no mediation ($a = 0.4$, $b = 0$).

n	Probability of outliers	Outlier shift d	OLS bootstrap	OLS Sobel	Box-Cox bootstrap	Winsorized bootstrap	Median bootstrap	ROBMED
100	0%	0	0.072	0.030	0.072	0.084	0.083	0.070
	1%	1	0.082	0.031	0.080	0.094	0.085	0.087
	1%	2	0.107	0.050	0.097	0.118	0.091	0.105
	1%	3	0.184	0.148	0.134	0.126	0.098	0.087
	1%	4	0.248	0.274	0.154	0.127	0.118	0.071
	1%	5	0.298	0.399	0.163	0.128	0.127	0.077
	1%	6	0.315	0.476	0.169	0.129	0.144	0.079
	2%	1	0.086	0.028	0.087	0.099	0.087	0.091
	2%	2	0.166	0.096	0.163	0.180	0.103	0.120
	2%	3	0.339	0.301	0.238	0.202	0.120	0.096
	2%	4	0.469	0.483	0.274	0.213	0.155	0.076
	2%	5	0.517	0.594	0.293	0.216	0.170	0.082
	2%	6	0.516	0.630	0.309	0.218	0.213	0.083
	3%	1	0.098	0.026	0.101	0.113	0.094	0.101
	3%	2	0.253	0.159	0.230	0.263	0.132	0.161
	3%	3	0.496	0.455	0.352	0.336	0.168	0.135
	3%	4	0.635	0.638	0.398	0.351	0.217	0.095
	3%	5	0.665	0.697	0.419	0.356	0.276	0.083
	3%	6	0.648	0.678	0.429	0.359	0.332	0.086
	4%	1	0.105	0.032	0.107	0.120	0.106	0.110
	4%	2	0.343	0.227	0.313	0.360	0.165	0.226
	4%	3	0.614	0.582	0.462	0.457	0.228	0.199
	4%	4	0.745	0.720	0.517	0.475	0.308	0.132
	4%	5	0.740	0.717	0.535	0.485	0.389	0.092
	4%	6	0.697	0.659	0.545	0.490	0.459	0.088
250	0%	0	0.068	0.045	0.066	0.077	0.073	0.071
	1%	1	0.065	0.046	0.061	0.080	0.071	0.072
	1%	2	0.131	0.102	0.122	0.129	0.086	0.091
	1%	3	0.285	0.314	0.209	0.138	0.112	0.072
	1%	4	0.455	0.582	0.260	0.141	0.125	0.064
	1%	5	0.575	0.750	0.281	0.143	0.141	0.069
	1%	6	0.662	0.835	0.301	0.144	0.180	0.069
	2%	1	0.078	0.050	0.077	0.088	0.086	0.093
	2%	2	0.279	0.256	0.256	0.286	0.125	0.163
	2%	3	0.598	0.684	0.451	0.325	0.180	0.092
	2%	4	0.821	0.896	0.523	0.336	0.265	0.061
	2%	5	0.890	0.961	0.561	0.340	0.326	0.070
	2%	6	0.921	0.977	0.586	0.342	0.402	0.070
	3%	1	0.090	0.060	0.096	0.107	0.103	0.121
	3%	2	0.436	0.412	0.402	0.443	0.193	0.256
	3%	3	0.843	0.887	0.653	0.513	0.301	0.107
	3%	4	0.952	0.985	0.747	0.528	0.422	0.058
	3%	5	0.977	0.991	0.787	0.533	0.531	0.067
	3%	6	0.983	0.992	0.807	0.537	0.641	0.068
	4%	1	0.113	0.071	0.113	0.146	0.131	0.154
	4%	2	0.614	0.608	0.569	0.624	0.279	0.364
	4%	3	0.951	0.972	0.846	0.718	0.455	0.154
	4%	4	0.986	0.997	0.902	0.732	0.600	0.071
	4%	5	0.992	0.993	0.923	0.741	0.723	0.064
	4%	6	0.970	0.983	0.929	0.743	0.803	0.068

Table 5. Bias and standard deviation (in parenthesis) for simulation design 3 with skew-t error distributions in the setting with mediation ($a = b = 0.4$). The parameters λ and ν control the skewness and kurtosis of the skew-t distribution.

n	λ	υ	Skewness	Excess kurtosis	OLS bootstrap	OLS Sobel	Box-Cox bootstrap	Winsorized bootstrap	Median bootstrap	ROBMED
100	$-\infty$	∞	-0.995	0.869	0.000 (0.041)	0.000 (0.041)	0.075 (0.080)	-0.002 (0.040)	0.002 (0.047)	0.002 (0.041)
	-2.174	∞	-0.5	0.347	0.000 (0.042)	0.000 (0.042)	0.052 (0.073)	-0.001 (0.043)	0.000 (0.047)	0.001 (0.045)
	0	∞	0	0	-0.001 (0.050)	-0.001 (0.050)	0.002 (0.065)	-0.001 (0.051)	0.000 (0.057)	0.000 (0.055)
	2.174	∞	0.5	0.347	-0.001 (0.042)	-0.001 (0.042)	0.038 (0.070)	-0.001 (0.042)	0.000 (0.046)	0.000 (0.044)
	∞	∞	0.995	0.869	-0.001 (0.040)	-0.001 (0.040)	0.054 (0.073)	-0.002 (0.040)	0.002 (0.046)	0.001 (0.040)
	$-\infty$	5	-2.55	20.109	0.000 (0.048)	0.000 (0.048)	0.084 (0.089)	-0.001 (0.043)	0.004 (0.044)	0.002 (0.038)
	-2.174	5	-1.869	14.138	-0.001 (0.050)	-0.001 (0.050)	0.056 (0.081)	-0.001 (0.046)	0.002 (0.046)	0.001 (0.044)
	0	5	0	6	-0.001 (0.058)	-0.001 (0.058)	0.002 (0.080)	-0.001 (0.056)	0.000 (0.056)	0.000 (0.055)
	2.174	5	1.869	14.138	-0.001 (0.050)	-0.001 (0.050)	0.007 (0.091)	-0.001 (0.045)	0.001 (0.045)	0.000 (0.043)
	∞	5	2.55	20.109	-0.001 (0.048)	-0.001 (0.048)	0.017 (0.098)	-0.002 (0.041)	0.003 (0.043)	0.001 (0.037)
	$-\infty$	2	-	-	-0.005 (0.135)	-0.005 (0.133)	0.131 (0.111)	0.001 (0.050)	0.005 (0.044)	0.001 (0.034)
	-2.174	2	-	-	-0.005 (0.136)	-0.005 (0.134)	0.099 (0.108)	0.000 (0.055)	0.003 (0.046)	0.001 (0.041)
	0	2	-	-	-0.003 (0.156)	-0.003 (0.155)	0.107 (0.139)	-0.001 (0.067)	0.000 (0.055)	0.001 (0.056)
	2.174	2	-	-	0.001 (0.157)	0.001 (0.155)	0.056 (0.163)	0.000 (0.053)	0.003 (0.046)	0.001 (0.042)
	∞	2	-	-	0.001 (0.154)	0.001 (0.153)	-0.006 (0.144)	0.000 (0.049)	0.005 (0.044)	0.002 (0.035)
250	$-\infty$	∞	-0.995	0.869	0.000 (0.024)	0.000 (0.024)	0.081 (0.051)	-0.002 (0.024)	0.000 (0.029)	0.000 (0.024)
	-2.174	∞	-0.5	0.347	0.000 (0.025)	0.000 (0.025)	0.056 (0.048)	0.000 (0.025)	0.000 (0.029)	0.000 (0.026)
	0	∞	0	0	0.000 (0.030)	0.000 (0.030)	0.002 (0.043)	0.000 (0.030)	0.000 (0.036)	0.000 (0.033)
	2.174	∞	0.5	0.347	0.000 (0.025)	0.000 (0.025)	0.033 (0.047)	0.000 (0.025)	0.000 (0.029)	0.000 (0.027)
	∞	∞	0.995	0.869	0.000 (0.024)	0.000 (0.024)	0.046 (0.051)	-0.002 (0.023)	0.001 (0.029)	0.000 (0.024)
	$-\infty$	5	-2.55	20.109	0.001 (0.030)	0.001 (0.030)	0.098 (0.060)	-0.002 (0.025)	0.002 (0.027)	0.001 (0.022)
	-2.174	5	-1.869	14.138	0.001 (0.031)	0.001 (0.031)	0.068 (0.058)	-0.001 (0.027)	0.001 (0.028)	0.000 (0.025)
	0	5	0	6	0.001 (0.036)	0.001 (0.036)	0.013 (0.064)	0.000 (0.034)	0.000 (0.035)	0.000 (0.033)
	2.174	5	1.869	14.138	0.000 (0.030)	0.000 (0.030)	0.001 (0.077)	0.000 (0.027)	0.001 (0.028)	0.000 (0.026)
	∞	5	2.55	20.109	0.000 (0.028)	0.000 (0.028)	0.003 (0.083)	-0.001 (0.025)	0.002 (0.027)	0.001 (0.022)
	$-\infty$	2	-	-	0.014 (0.290)	0.014 (0.294)	0.170 (0.081)	0.000 (0.029)	0.003 (0.027)	0.001 (0.020)
	-2.174	2	-	-	0.013 (0.267)	0.014 (0.271)	0.134 (0.081)	0.000 (0.032)	0.002 (0.029)	0.001 (0.025)
	0	2	-	-	0.007 (0.135)	0.007 (0.135)	0.165 (0.109)	0.000 (0.041)	0.001 (0.034)	0.000 (0.034)
	2.174	2	-	-	0.001 (0.068)	0.001 (0.068)	0.110 (0.143)	0.001 (0.033)	0.002 (0.028)	0.001 (0.024)
	∞	2	-	-	0.000 (0.066)	0.000 (0.066)	0.015 (0.119)	0.001 (0.029)	0.003 (0.027)	0.001 (0.020)

Table 6. Rate of rejection with correct sign of the indirect effect (realized power; the higher the better) for simulation design 3 with skew-t error distributions in the setting with mediation ($a = b = 0.4$). The parameters λ and ν control the skewness and kurtosis of the skew-t distribution.

n	λ	υ	Skewness	Excess kurtosis	OLS bootstrap	OLS Sobel	Box-Cox bootstrap	Winsorized bootstrap	Median bootstrap	ROBMED
100	$-\infty$	∞	-0.995	0.869	0.999	1	0.999	1	0.965	1
	-2.174	∞	-0.5	0.347	0.999	0.999	0.998	0.998	0.947	0.995
	0	∞	0	0	0.989	0.973	0.987	0.989	0.873	0.968
	2.174	∞	0.5	0.347	0.999	0.996	0.997	0.997	0.963	0.993
	∞	∞	0.995	0.869	0.996	0.996	0.996	0.997	0.966	0.998
	$-\infty$	5	-2.55	20.109	0.970	0.976	0.996	0.999	0.983	0.998
	-2.174	5	-1.869	14.138	0.970	0.968	0.994	0.995	0.968	0.996
	0	5	0	6	0.907	0.860	0.934	0.957	0.847	0.936
	2.174	5	1.869	14.138	0.965	0.968	0.988	0.993	0.970	0.995
	∞	5	2.55	20.109	0.970	0.976	0.991	0.996	0.982	0.998
	$-\infty$	2	-	-	0.634	0.453	0.986	0.969	0.950	0.994
	-2.174	2	-	-	0.625	0.428	0.951	0.942	0.918	0.977
	0	2	-	-	0.542	0.330	0.773	0.812	0.779	0.865
	2.174	2	-	-	0.639	0.450	0.940	0.937	0.906	0.980
	∞	2	-	-	0.653	0.483	0.983	0.959	0.942	0.998
250	$-\infty$	∞	-0.995	0.869	1	1	1	1	1	1
	-2.174	∞	-0.5	0.347	1	1	1	1	1	1
	0	∞	0	0	1	1	1	1	1	1
	2.174	∞	0.5	0.347	1	1	1	1	1	1
	∞	∞	0.995	0.869	1	1	1	1	1	1
	$-\infty$	5	-2.55	20.109	0.999	1	1	1	1	1
	-2.174	5	-1.869	14.138	0.999	1	1	1	1	1
	0	5	0	6	0.996	0.998	0.999	0.999	0.999	1
	2.174	5	1.869	14.138	0.999	1	1	1	1	1
	∞	5	2.55	20.109	0.999	1	1	1	1	1
	$-\infty$	2	-	-	0.821	0.793	1	1	1	1
	-2.174	2	-	-	0.813	0.779	0.998	0.998	0.999	1
	0	2	-	-	0.776	0.696	0.987	0.990	0.991	0.997
	2.174	2	-	-	0.820	0.783	1	0.998	0.999	1
	∞	2	-	-	0.817	0.793	1	1	0.999	1

Table 7. Bias and standard deviation (in parenthesis) for simulation design 3 with skew-t error distributions in the setting with no mediation ($a = 0.4$, $b = 0$). The parameters λ and ν control the skewness and kurtosis of the skew-t distribution.

n	λ	ν	Skewness	Excess kurtosis	OLS bootstrap	OLS Sobel	Box-Cox bootstrap	Winsorized bootstrap	Median bootstrap	ROBMED
100	$-\infty$	∞	-0.995	0.869	0.001 (0.042)	0.001 (0.042)	0.004 (0.068)	-0.001 (0.042)	0.003 (0.048)	0.002 (0.042)
	-2.174	∞	-0.5	0.347	0.000 (0.042)	0.000 (0.042)	0.002 (0.061)	0.000 (0.042)	0.001 (0.046)	0.001 (0.044)
	0	∞	0	0	0.000 (0.042)	0.000 (0.042)	0.001 (0.044)	0.000 (0.043)	0.000 (0.047)	0.001 (0.046)
	2.174	∞	0.5	0.347	0.000 (0.042)	0.000 (0.042)	0.001 (0.055)	0.000 (0.042)	0.001 (0.045)	0.000 (0.044)
	∞	∞	0.995	0.869	0.000 (0.041)	0.000 (0.041)	0.002 (0.058)	-0.002 (0.041)	0.002 (0.046)	0.001 (0.041)
	$-\infty$	5	-2.55	20.109	0.001 (0.045)	0.001 (0.045)	0.004 (0.064)	-0.001 (0.040)	0.005 (0.040)	0.003 (0.035)
	-2.174	5	-1.869	14.138	0.000 (0.044)	0.001 (0.044)	0.002 (0.060)	0.000 (0.041)	0.002 (0.039)	0.002 (0.038)
	0	5	0	6	0.000 (0.044)	0.000 (0.044)	0.001 (0.051)	0.000 (0.042)	0.000 (0.040)	0.001 (0.040)
	2.174	5	1.869	14.138	0.000 (0.043)	0.000 (0.043)	0.001 (0.052)	0.000 (0.040)	0.002 (0.038)	0.001 (0.037)
	∞	5	2.55	20.109	0.000 (0.043)	0.000 (0.043)	0.001 (0.042)	-0.001 (0.039)	0.004 (0.038)	0.001 (0.033)
	$-\infty$	2	-	-	0.000 (0.081)	0.000 (0.075)	0.004 (0.067)	0.002 (0.038)	0.007 (0.028)	0.002 (0.022)
	-2.174	2	-	-	0.000 (0.080)	0.000 (0.075)	0.003 (0.066)	0.001 (0.039)	0.005 (0.028)	0.002 (0.025)
	0	2	-	-	0.000 (0.097)	0.001 (0.095)	0.001 (0.079)	0.001 (0.040)	0.001 (0.028)	0.001 (0.029)
	2.174	2	-	-	0.001 (0.080)	0.001 (0.079)	0.002 (0.077)	0.001 (0.036)	0.004 (0.027)	0.001 (0.025)
	∞	2	-	-	0.001 (0.073)	0.001 (0.072)	0.000 (0.039)	0.001 (0.035)	0.007 (0.027)	0.002 (0.022)
250	$-\infty$	∞	-0.995	0.869	0.000 (0.025)	0.000 (0.025)	0.002 (0.043)	-0.002 (0.025)	0.000 (0.029)	-0.001 (0.025)
	-2.174	∞	-0.5	0.347	0.000 (0.025)	0.000 (0.025)	0.001 (0.038)	-0.001 (0.025)	-0.001 (0.027)	-0.001 (0.026)
	0	∞	0	0	0.000 (0.025)	0.000 (0.025)	0.000 (0.026)	0.000 (0.025)	-0.001 (0.028)	-0.001 (0.026)
	2.174	∞	0.5	0.347	0.000 (0.025)	0.000 (0.025)	0.001 (0.031)	-0.001 (0.025)	0.000 (0.028)	0.000 (0.026)
	∞	∞	0.995	0.869	0.000 (0.025)	0.000 (0.025)	0.002 (0.032)	-0.002 (0.025)	0.000 (0.029)	0.000 (0.025)
	$-\infty$	5	-2.55	20.109	0.001 (0.028)	0.001 (0.028)	0.003 (0.041)	-0.002 (0.024)	0.001 (0.023)	0.000 (0.019)
	-2.174	5	-1.869	14.138	0.001 (0.028)	0.001 (0.028)	0.001 (0.038)	-0.001 (0.024)	0.000 (0.023)	-0.001 (0.021)
	0	5	0	6	0.000 (0.026)	0.001 (0.026)	0.000 (0.034)	0.000 (0.024)	-0.001 (0.023)	-0.001 (0.023)
	2.174	5	1.869	14.138	0.000 (0.026)	0.000 (0.026)	0.001 (0.032)	-0.001 (0.024)	0.000 (0.023)	0.000 (0.022)
	∞	5	2.55	20.109	0.000 (0.026)	0.000 (0.026)	0.002 (0.021)	-0.002 (0.024)	0.002 (0.023)	0.001 (0.019)
	$-\infty$	2	-	-	0.013 (0.337)	0.013 (0.342)	0.002 (0.045)	0.000 (0.021)	0.003 (0.014)	0.001 (0.011)
	-2.174	2	-	-	0.012 (0.306)	0.012 (0.312)	0.001 (0.043)	0.000 (0.022)	0.002 (0.014)	0.000 (0.013)
	0	2	-	-	0.005 (0.130)	0.005 (0.130)	0.001 (0.053)	0.000 (0.022)	0.000 (0.014)	-0.001 (0.015)
	2.174	2	-	-	0.000 (0.040)	0.000 (0.038)	0.002 (0.054)	0.001 (0.022)	0.002 (0.014)	0.001 (0.012)
	∞	2	-	-	-0.001 (0.039)	-0.001 (0.038)	0.002 (0.025)	0.001 (0.021)	0.003 (0.014)	0.001 (0.010)

Table 8. Rejection rate (realized size of the tests; the closer to $\alpha = 0.05$ the better) for simulation design 3 with skew-t error distributions in the setting with no mediation ($a = 0.4$, $b = 0$). The parameters λ and ν control the skewness and kurtosis of the skew-t distribution.

n	λ	υ	Skewness	Excess kurtosis	OLS bootstrap	OLS Sobel	Box-Cox bootstrap	Winsorized bootstrap	Median bootstrap	ROBMED
100	$-\infty$	∞	-0.995	0.869	0.082	0.052	0.071	0.076	0.084	0.074
	-2.174	∞	-0.5	0.347	0.076	0.044	0.074	0.091	0.084	0.072
	0	∞	0	0	0.086	0.027	0.082	0.098	0.081	0.081
	2.174	∞	0.5	0.347	0.080	0.041	0.071	0.085	0.075	0.077
	∞	∞	0.995	0.869	0.066	0.035	0.063	0.071	0.071	0.069
	$-\infty$	5	-2.55	20.109	0.099	0.044	0.080	0.091	0.094	0.082
	-2.174	5	-1.869	14.138	0.099	0.042	0.085	0.093	0.090	0.089
	0	5	0	6	0.091	0.017	0.083	0.101	0.075	0.077
	2.174	5	1.869	14.138	0.093	0.028	0.076	0.087	0.072	0.079
	∞	5	2.55	20.109	0.083	0.030	0.072	0.081	0.084	0.077
	$-\infty$	2	-	-	0.113	0.013	0.086	0.098	0.096	0.107
	-2.174	2	-	-	0.105	0.014	0.096	0.107	0.086	0.105
	0	2	-	-	0.064	0.007	0.071	0.092	0.064	0.086
	2.174	2	-	-	0.086	0.006	0.079	0.087	0.077	0.104
	∞	2	-	-	0.096	0.011	0.075	0.080	0.073	0.104
250	$-\infty$	∞	-0.995	0.869	0.060	0.046	0.060	0.055	0.053	0.052
	-2.174	∞	-0.5	0.347	0.058	0.049	0.056	0.061	0.044	0.050
	0	∞	0	0	0.058	0.037	0.058	0.064	0.059	0.056
	2.174	∞	0.5	0.347	0.057	0.041	0.045	0.061	0.037	0.046
	∞	∞	0.995	0.869	0.051	0.043	0.055	0.062	0.050	0.047
	$-\infty$	5	-2.55	20.109	0.074	0.051	0.069	0.065	0.066	0.060
	-2.174	5	-1.869	14.138	0.079	0.045	0.066	0.077	0.063	0.061
	0	5	0	6	0.076	0.035	0.069	0.079	0.062	0.059
	2.174	5	1.869	14.138	0.082	0.041	0.052	0.072	0.053	0.056
	∞	5	2.55	20.109	0.073	0.038	0.056	0.068	0.057	0.055
	$-\infty$	2	-	-	0.097	0.024	0.075	0.070	0.063	0.079
	-2.174	2	-	-	0.094	0.022	0.076	0.087	0.073	0.078
	0	2	-	-	0.112	0.014	0.074	0.095	0.056	0.074
	2.174	2	-	-	0.103	0.023	0.067	0.079	0.070	0.075
	∞	2	-	-	0.101	0.023	0.058	0.081	0.067	0.071

Table 9. Bias and standard deviation (in parenthesis) for simulation design 3 with non-normal errors via Fleishman's method, in the setting with mediation ($a = b = 0.4$).

n	Skewness	Excess kurtosis	OLS bootstrap	OLS Sobel	Box-Cox bootstrap	Winsorized bootstrap	Median bootstrap	ROBMED
100	-1	1	-0.002 (0.050)	-0.002 (0.050)	0.037 (0.069)	-0.003 (0.048)	-0.002 (0.056)	-0.003 (0.049)
	-1	2	-0.002 (0.050)	-0.002 (0.050)	0.033 (0.069)	-0.003 (0.047)	-0.002 (0.050)	-0.003 (0.047)
	-1	3	-0.002 (0.050)	-0.002 (0.050)	0.034 (0.073)	-0.003 (0.047)	-0.002 (0.047)	-0.003 (0.046)
	-1	4	-0.002 (0.051)	-0.002 (0.051)	0.039 (0.079)	-0.003 (0.046)	-0.002 (0.044)	-0.003 (0.044)
	-0.5	0	-0.002 (0.049)	-0.002 (0.049)	0.017 (0.064)	-0.003 (0.049)	-0.003 (0.057)	-0.003 (0.053)
	-0.5	1	-0.002 (0.050)	-0.002 (0.050)	0.017 (0.066)	-0.003 (0.048)	-0.003 (0.051)	-0.004 (0.050)
	-0.5	2	-0.002 (0.050)	-0.002 (0.050)	0.022 (0.072)	-0.003 (0.047)	-0.003 (0.047)	-0.004 (0.047)
	-0.5	3	-0.002 (0.050)	-0.002 (0.050)	0.030 (0.079)	-0.003 (0.047)	-0.003 (0.045)	-0.004 (0.045)
	-0.5	4	-0.002 (0.050)	-0.002 (0.050)	0.038 (0.086)	-0.003 (0.046)	-0.003 (0.043)	-0.004 (0.044)
	0	-1	-0.003 (0.048)	-0.003 (0.049)	0.001 (0.060)	-0.003 (0.049)	-0.004 (0.073)	-0.003 (0.058)
	0	0	-0.003 (0.049)	-0.003 (0.049)	0.001 (0.062)	-0.003 (0.049)	-0.004 (0.055)	-0.004 (0.053)
	0	1	-0.002 (0.049)	-0.002 (0.049)	0.007 (0.068)	-0.003 (0.048)	-0.004 (0.050)	-0.004 (0.049)
	0	2	-0.002 (0.050)	-0.002 (0.050)	0.015 (0.075)	-0.003 (0.048)	-0.003 (0.047)	-0.004 (0.047)
	0	3	-0.002 (0.050)	-0.002 (0.050)	0.025 (0.083)	-0.003 (0.047)	-0.003 (0.044)	-0.004 (0.045)
	0	4	-0.002 (0.050)	-0.002 (0.050)	0.035 (0.090)	-0.003 (0.046)	-0.003 (0.042)	-0.004 (0.043)
	0.5	0	-0.003 (0.049)	-0.003 (0.048)	-0.014 (0.059)	-0.004 (0.049)	-0.003 (0.056)	-0.003 (0.053)
	0.5	1	-0.003 (0.049)	-0.003 (0.049)	-0.007 (0.066)	-0.003 (0.048)	-0.003 (0.051)	-0.003 (0.050)
	0.5	2	-0.003 (0.049)	-0.003 (0.049)	0.004 (0.074)	-0.003 (0.048)	-0.003 (0.047)	-0.003 (0.047)
	0.5	3	-0.002 (0.050)	-0.003 (0.050)	0.015 (0.082)	-0.003 (0.047)	-0.003 (0.044)	-0.003 (0.045)
	0.5	4	-0.002 (0.050)	-0.002 (0.050)	0.026 (0.091)	-0.003 (0.046)	-0.003 (0.042)	-0.003 (0.043)
	1	1	-0.003 (0.049)	-0.003 (0.049)	-0.024 (0.063)	-0.005 (0.047)	-0.002 (0.054)	-0.002 (0.049)
	1	2	-0.003 (0.049)	-0.003 (0.049)	-0.016 (0.069)	-0.004 (0.047)	-0.002 (0.049)	-0.002 (0.047)
	1	3	-0.003 (0.049)	-0.003 (0.049)	-0.003 (0.079)	-0.003 (0.047)	-0.002 (0.046)	-0.002 (0.045)
	1	4	-0.003 (0.050)	-0.003 (0.050)	0.009 (0.087)	-0.003 (0.046)	-0.002 (0.043)	-0.003 (0.043)
250	-1	1	0.001 (0.031)	0.001 (0.031)	0.055 (0.051)	0.000 (0.030)	0.002 (0.036)	0.001 (0.031)
	-1	2	0.001 (0.031)	0.001 (0.031)	0.048 (0.050)	0.000 (0.029)	0.002 (0.032)	0.001 (0.030)
	-1	3	0.001 (0.031)	0.001 (0.031)	0.047 (0.053)	0.001 (0.029)	0.002 (0.030)	0.001 (0.028)
	-1	4	0.001 (0.031)	0.001 (0.031)	0.054 (0.058)	0.001 (0.029)	0.001 (0.028)	0.001 (0.027)
	-0.5	0	0.001 (0.030)	0.001 (0.030)	0.029 (0.046)	0.001 (0.031)	0.002 (0.037)	0.001 (0.033)
	-0.5	1	0.001 (0.030)	0.001 (0.030)	0.025 (0.048)	0.001 (0.030)	0.001 (0.033)	0.001 (0.031)
	-0.5	2	0.001 (0.030)	0.001 (0.030)	0.031 (0.053)	0.001 (0.030)	0.001 (0.030)	0.001 (0.029)
	-0.5	3	0.001 (0.030)	0.001 (0.030)	0.043 (0.060)	0.001 (0.029)	0.001 (0.029)	0.001 (0.028)
	-0.5	4	0.001 (0.030)	0.001 (0.030)	0.056 (0.066)	0.001 (0.029)	0.001 (0.027)	0.001 (0.027)
	0	-1	0.001 (0.031)	0.001 (0.031)	0.009 (0.044)	0.001 (0.031)	0.002 (0.050)	0.001 (0.037)
	0	0	0.001 (0.030)	0.001 (0.030)	0.003 (0.043)	0.001 (0.031)	0.001 (0.036)	0.001 (0.033)
	0	1	0.001 (0.030)	0.001 (0.030)	0.008 (0.049)	0.001 (0.030)	0.001 (0.032)	0.001 (0.031)
	0	2	0.001 (0.030)	0.001 (0.030)	0.022 (0.058)	0.001 (0.029)	0.001 (0.030)	0.001 (0.029)
	0	3	0.001 (0.030)	0.001 (0.030)	0.038 (0.065)	0.001 (0.029)	0.001 (0.028)	0.000 (0.028)
	0	4	0.001 (0.030)	0.001 (0.030)	0.055 (0.072)	0.001 (0.029)	0.001 (0.027)	0.000 (0.027)
	0.5	0	0.001 (0.030)	0.001 (0.030)	-0.019 (0.041)	0.000 (0.031)	0.001 (0.037)	0.001 (0.033)
	0.5	1	0.001 (0.030)	0.001 (0.030)	-0.014 (0.048)	0.001 (0.030)	0.001 (0.033)	0.001 (0.031)
	0.5	2	0.001 (0.030)	0.001 (0.030)	0.003 (0.058)	0.001 (0.030)	0.001 (0.030)	0.001 (0.029)
	0.5	3	0.001 (0.030)	0.001 (0.030)	0.023 (0.068)	0.001 (0.029)	0.001 (0.028)	0.001 (0.028)
	0.5	4	0.001 (0.030)	0.001 (0.030)	0.043 (0.075)	0.001 (0.029)	0.001 (0.027)	0.001 (0.027)
	1	1	0.001 (0.030)	0.000 (0.030)	-0.038 (0.043)	-0.001 (0.030)	0.002 (0.036)	0.001 (0.031)
	1	2	0.001 (0.030)	0.000 (0.030)	-0.031 (0.051)	0.000 (0.030)	0.001 (0.032)	0.001 (0.030)
	1	3	0.001 (0.030)	0.000 (0.030)	-0.011 (0.062)	0.001 (0.029)	0.001 (0.030)	0.001 (0.028)
	1	4	0.001 (0.030)	0.000 (0.030)	0.014 (0.074)	0.001 (0.029)	0.001 (0.028)	0.001 (0.027)

Table 10. Rate of rejection with correct sign of the indirect effect (realized power; the higher the better) for simulation design 3 with non-normal errors via Fleishman's method, in the setting with mediation ($a = b = 0.4$).

n	Skewness	Excess kurtosis	OLS bootstrap	OLS Sobel	Box-Cox bootstrap	Winsorized bootstrap	Median bootstrap	ROBMED
100	-1	1	0.988	0.977	0.993	0.999	0.879	0.984
	-1	2	0.983	0.971	0.991	0.996	0.919	0.984
	-1	3	0.973	0.965	0.989	0.995	0.947	0.989
	-1	4	0.966	0.962	0.985	0.995	0.959	0.992
	-0.5	0	0.994	0.972	0.993	0.996	0.844	0.966
	-0.5	1	0.987	0.967	0.991	0.995	0.902	0.974
	-0.5	2	0.977	0.964	0.987	0.995	0.940	0.977
	-0.5	3	0.970	0.962	0.986	0.994	0.955	0.985
	-0.5	4	0.963	0.958	0.986	0.994	0.966	0.989
	0	-1	0.993	0.978	0.989	0.993	0.587	0.960
	0	0	0.991	0.964	0.989	0.992	0.867	0.965
	0	1	0.986	0.963	0.988	0.993	0.906	0.973
	0	2	0.979	0.959	0.986	0.995	0.935	0.978
	0	3	0.968	0.958	0.985	0.994	0.955	0.982
	0	4	0.958	0.959	0.983	0.995	0.963	0.987
	0.5	0	0.991	0.968	0.994	0.991	0.846	0.974
	0.5	1	0.984	0.964	0.989	0.991	0.903	0.974
	0.5	2	0.975	0.961	0.987	0.993	0.934	0.978
	0.5	3	0.966	0.958	0.986	0.994	0.949	0.986
	0.5	4	0.962	0.957	0.983	0.995	0.963	0.988
	1	1	0.985	0.966	0.996	0.992	0.877	0.987
	1	2	0.975	0.960	0.996	0.992	0.922	0.983
	1	3	0.968	0.959	0.991	0.992	0.941	0.989
	1	4	0.965	0.958	0.988	0.994	0.953	0.991
250	-1	1	1	1	1	1	1	1
	-1	2	1	1	1	1	1	1
	-1	3	1	1	1	1	1	1
	-1	4	1	1	1	1	1	1
	-0.5	0	1	1	1	1	0.999	1
	-0.5	1	1	1	1	1	1	1
	-0.5	2	1	1	1	1	1	1
	-0.5	3	0.999	1	1	1	1	1
	-0.5	4	0.998	1	1	1	1	1
	0	-1	1	1	1	1	0.979	1
	0	0	1	1	1	1	0.999	1
	0	1	1	1	1	1	1	1
	0	2	1	1	1	1	1	1
	0	3	0.999	1	1	1	1	1
	0	4	0.998	1	1	1	1	1
	0.5	0	1	1	1	1	0.999	1
	0.5	1	1	1	1	1	1	1
	0.5	2	1	1	1	1	1	1
	0.5	3	0.999	1	1	1	1	1
	0.5	4	0.998	1	1	1	1	1
	1	1	1	1	1	1	1	1
	1	2	1	1	1	1	1	1
	1	3	1	1	1	1	1	1
	1	4	0.999	1	1	1	1	1

Table 11. Bias and standard deviation (in parenthesis) for simulation design 3 with non-normal errors via Fleishman's method, in the setting with no mediation ($a = 0.4$, $b = 0$).

n	Skewness	Excess kurtosis	OLS bootstrap	OLS Sobel	Box-Cox bootstrap	Winsorized bootstrap	Median bootstrap	ROBMED
100	-1	1	-0.001 (0.043)	-0.002 (0.043)	-0.002 (0.056)	-0.003 (0.042)	0.000 (0.048)	-0.001 (0.042)
	-1	2	-0.002 (0.043)	-0.002 (0.043)	-0.002 (0.054)	-0.002 (0.041)	-0.001 (0.043)	-0.002 (0.040)
	-1	3	-0.002 (0.043)	-0.002 (0.043)	-0.002 (0.055)	-0.002 (0.041)	-0.001 (0.040)	-0.002 (0.039)
	-1	4	-0.001 (0.043)	-0.002 (0.043)	-0.002 (0.060)	-0.002 (0.040)	-0.001 (0.038)	-0.002 (0.038)
	-0.5	0	-0.002 (0.042)	-0.002 (0.042)	-0.002 (0.050)	-0.002 (0.042)	-0.002 (0.049)	-0.002 (0.045)
	-0.5	1	-0.002 (0.042)	-0.002 (0.042)	-0.002 (0.049)	-0.002 (0.041)	-0.002 (0.044)	-0.002 (0.043)
	-0.5	2	-0.002 (0.042)	-0.002 (0.042)	-0.002 (0.054)	-0.002 (0.041)	-0.002 (0.041)	-0.002 (0.041)
	-0.5	3	-0.002 (0.042)	-0.002 (0.042)	-0.003 (0.060)	-0.002 (0.041)	-0.002 (0.039)	-0.002 (0.039)
	-0.5	4	-0.002 (0.043)	-0.002 (0.043)	-0.003 (0.066)	-0.002 (0.040)	-0.002 (0.037)	-0.002 (0.038)
	0	-1	-0.002 (0.041)	-0.002 (0.041)	-0.002 (0.049)	-0.002 (0.041)	-0.002 (0.064)	-0.002 (0.049)
	0	0	-0.002 (0.041)	-0.002 (0.041)	-0.002 (0.043)	-0.002 (0.042)	-0.002 (0.047)	-0.003 (0.045)
	0	1	-0.002 (0.041)	-0.002 (0.041)	-0.002 (0.048)	-0.002 (0.041)	-0.002 (0.043)	-0.003 (0.043)
	0	2	-0.002 (0.042)	-0.002 (0.042)	-0.003 (0.056)	-0.002 (0.041)	-0.002 (0.040)	-0.003 (0.041)
	0	3	-0.002 (0.042)	-0.002 (0.042)	-0.003 (0.063)	-0.002 (0.041)	-0.002 (0.038)	-0.003 (0.039)
	0	4	-0.002 (0.042)	-0.002 (0.042)	-0.003 (0.069)	-0.002 (0.040)	-0.002 (0.036)	-0.003 (0.038)
	0.5	0	-0.002 (0.041)	-0.002 (0.041)	-0.002 (0.036)	-0.003 (0.041)	-0.002 (0.048)	-0.002 (0.046)
	0.5	1	-0.002 (0.041)	-0.002 (0.041)	-0.002 (0.041)	-0.002 (0.041)	-0.002 (0.043)	-0.002 (0.043)
	0.5	2	-0.002 (0.041)	-0.002 (0.041)	-0.002 (0.052)	-0.002 (0.041)	-0.002 (0.040)	-0.002 (0.041)
	0.5	3	-0.002 (0.042)	-0.002 (0.042)	-0.003 (0.061)	-0.002 (0.041)	-0.002 (0.038)	-0.002 (0.040)
	0.5	4	-0.002 (0.042)	-0.002 (0.042)	-0.003 (0.069)	-0.002 (0.041)	-0.002 (0.036)	-0.002 (0.038)
	1	1	-0.002 (0.041)	-0.002 (0.041)	-0.001 (0.029)	-0.004 (0.041)	0.000 (0.045)	-0.001 (0.042)
	1	2	-0.002 (0.041)	-0.002 (0.041)	-0.001 (0.036)	-0.003 (0.041)	-0.001 (0.042)	-0.002 (0.041)
	1	3	-0.002 (0.041)	-0.002 (0.041)	-0.002 (0.049)	-0.003 (0.041)	-0.001 (0.039)	-0.002 (0.040)
	1	4	-0.002 (0.042)	-0.002 (0.041)	-0.003 (0.060)	-0.002 (0.041)	-0.001 (0.037)	-0.002 (0.038)
250	-1	1	0.001 (0.026)	0.001 (0.026)	0.001 (0.038)	-0.001 (0.026)	0.001 (0.030)	0.001 (0.025)
	-1	2	0.000 (0.026)	0.000 (0.026)	0.001 (0.035)	-0.001 (0.025)	0.001 (0.027)	0.000 (0.024)
	-1	3	0.000 (0.026)	0.000 (0.026)	0.001 (0.036)	0.000 (0.025)	0.001 (0.025)	0.000 (0.023)
	-1	4	0.000 (0.026)	0.000 (0.026)	0.001 (0.040)	0.000 (0.025)	0.001 (0.023)	0.000 (0.023)
	-0.5	0	0.000 (0.026)	0.000 (0.026)	0.001 (0.034)	0.000 (0.026)	0.000 (0.031)	0.000 (0.028)
	-0.5	1	0.000 (0.026)	0.000 (0.026)	0.001 (0.032)	0.000 (0.026)	0.000 (0.028)	0.000 (0.026)
	-0.5	2	0.000 (0.026)	0.000 (0.026)	0.001 (0.035)	0.000 (0.025)	0.000 (0.025)	0.000 (0.025)
	-0.5	3	0.000 (0.026)	0.000 (0.026)	0.001 (0.040)	0.000 (0.025)	0.000 (0.024)	0.000 (0.023)
	-0.5	4	0.000 (0.026)	0.000 (0.026)	0.001 (0.045)	0.000 (0.025)	0.000 (0.023)	0.000 (0.022)
	0	-1	0.000 (0.026)	0.000 (0.026)	0.000 (0.033)	0.000 (0.026)	0.000 (0.042)	0.001 (0.032)
	0	0	0.000 (0.026)	0.000 (0.026)	0.000 (0.027)	0.000 (0.026)	0.000 (0.030)	0.000 (0.028)
	0	1	0.000 (0.026)	0.000 (0.026)	0.001 (0.030)	0.000 (0.026)	0.000 (0.027)	0.000 (0.026)
	0	2	0.000 (0.026)	0.000 (0.026)	0.001 (0.037)	0.000 (0.026)	0.000 (0.025)	0.000 (0.024)
	0	3	0.000 (0.026)	0.000 (0.026)	0.001 (0.042)	0.000 (0.025)	0.000 (0.024)	0.000 (0.023)
	0	4	0.000 (0.026)	0.000 (0.026)	0.001 (0.047)	0.000 (0.025)	0.000 (0.022)	0.000 (0.022)
	0.5	0	0.000 (0.026)	0.000 (0.026)	0.000 (0.021)	0.000 (0.026)	0.000 (0.031)	0.000 (0.028)
	0.5	1	0.000 (0.025)	0.000 (0.025)	0.000 (0.024)	0.000 (0.026)	0.000 (0.028)	0.000 (0.026)
	0.5	2	0.000 (0.025)	0.000 (0.025)	0.001 (0.034)	0.000 (0.026)	0.000 (0.025)	0.000 (0.025)
	0.5	3	0.000 (0.025)	0.000 (0.025)	0.001 (0.041)	0.000 (0.025)	0.000 (0.024)	0.000 (0.023)
	0.5	4	0.000 (0.026)	0.000 (0.025)	0.001 (0.047)	0.000 (0.025)	0.000 (0.022)	0.000 (0.022)
	1	1	0.000 (0.026)	0.000 (0.026)	0.000 (0.015)	-0.002 (0.026)	0.001 (0.030)	0.000 (0.026)
	1	2	0.000 (0.025)	0.000 (0.025)	0.000 (0.018)	-0.001 (0.026)	0.000 (0.027)	0.000 (0.025)
	1	3	0.000 (0.025)	0.000 (0.025)	0.001 (0.031)	0.000 (0.026)	0.000 (0.025)	0.000 (0.024)
	1	4	0.000 (0.025)	0.000 (0.025)	0.001 (0.041)	0.000 (0.025)	0.000 (0.023)	0.000 (0.023)

Table 12. Rejection rate (realized size of the tests; the closer to $\alpha = 0.05$ the better) for simulation design 3 with non-normal errors via Fleishman's method, in the setting with no mediation ($a = 0.4$, $b = 0$).

n	Skewness	Excess kurtosis	OLS bootstrap	OLS Sobel	Box-Cox bootstrap	Winsorized bootstrap	Median bootstrap	ROBMED
100	-1	1	0.087	0.028	0.080	0.094	0.075	0.078
	-1	2	0.089	0.025	0.077	0.090	0.069	0.072
	-1	3	0.093	0.025	0.076	0.090	0.070	0.077
	-1	4	0.094	0.028	0.077	0.090	0.067	0.080
	-0.5	0	0.082	0.026	0.074	0.085	0.069	0.074
	-0.5	1	0.084	0.024	0.080	0.094	0.069	0.082
	-0.5	2	0.094	0.026	0.079	0.094	0.074	0.079
	-0.5	3	0.093	0.026	0.081	0.095	0.074	0.085
	-0.5	4	0.098	0.025	0.081	0.096	0.073	0.087
	0	-1	0.073	0.025	0.074	0.069	0.084	0.082
	0	0	0.084	0.023	0.077	0.094	0.073	0.084
	0	1	0.089	0.024	0.082	0.101	0.071	0.085
	0	2	0.093	0.027	0.083	0.100	0.075	0.084
	0	3	0.095	0.027	0.083	0.099	0.077	0.087
	0	4	0.097	0.027	0.082	0.099	0.070	0.089
	0.5	0	0.084	0.022	0.076	0.090	0.076	0.085
	0.5	1	0.086	0.023	0.081	0.097	0.073	0.090
	0.5	2	0.089	0.026	0.079	0.095	0.081	0.089
	0.5	3	0.088	0.027	0.081	0.098	0.084	0.088
	0.5	4	0.090	0.029	0.079	0.098	0.080	0.089
	1	1	0.085	0.022	0.070	0.089	0.084	0.087
	1	2	0.090	0.024	0.078	0.095	0.089	0.091
	1	3	0.089	0.025	0.080	0.099	0.085	0.087
	1	4	0.090	0.025	0.079	0.099	0.082	0.086
250	-1	1	0.067	0.044	0.060	0.073	0.068	0.060
	-1	2	0.057	0.041	0.053	0.071	0.054	0.056
	-1	3	0.063	0.038	0.056	0.065	0.055	0.055
	-1	4	0.061	0.039	0.059	0.065	0.053	0.056
	-0.5	0	0.056	0.036	0.055	0.066	0.059	0.058
	-0.5	1	0.054	0.034	0.054	0.069	0.049	0.050
	-0.5	2	0.059	0.035	0.055	0.070	0.057	0.054
	-0.5	3	0.060	0.035	0.053	0.069	0.058	0.057
	-0.5	4	0.062	0.036	0.058	0.070	0.059	0.060
	0	-1	0.068	0.052	0.067	0.070	0.080	0.068
	0	0	0.054	0.038	0.052	0.066	0.055	0.061
	0	1	0.058	0.034	0.050	0.074	0.058	0.059
	0	2	0.059	0.032	0.056	0.075	0.051	0.059
	0	3	0.062	0.036	0.055	0.075	0.050	0.061
	0	4	0.064	0.037	0.056	0.077	0.050	0.059
	0.5	0	0.053	0.036	0.054	0.067	0.070	0.066
	0.5	1	0.060	0.037	0.051	0.072	0.063	0.061
	0.5	2	0.060	0.036	0.053	0.076	0.055	0.060
	0.5	3	0.061	0.035	0.054	0.078	0.052	0.058
	0.5	4	0.063	0.036	0.054	0.080	0.056	0.058
	1	1	0.067	0.033	0.058	0.082	0.066	0.060
	1	2	0.064	0.034	0.056	0.074	0.066	0.064
	1	3	0.065	0.035	0.057	0.076	0.061	0.062
	1	4	0.069	0.035	0.057	0.078	0.065	0.062

Table 13. Literature review.

No	Journal	Year	Reference	OLS	Outliers	Model assumptions	Bootstrap	PROCESS macro
1	AMJ	2019	Clarke, J. S., Cornelissen, J. P., & Healey, M. P. (2019). Actions speak louder than words: How figurative language and gesturing in entrepreneurial pitches influences investment judgments. Academy of Management Journal, 62(2), 335-360.	Yes	No	No	Yes	Yes
2	AMJ	2019	Lin, S. H., Scott, B. A., & Matta, F. K. (2019). The dark side of transformational leader behaviors for leaders themselves: A conservation of resources perspective. Academy of Management Journal, 62(5), 1556-1582.	No	No	No	Yes	No
3	AMJ	2019	Mitchell, M. S., Greenbaum, R. L., Vogel, R. M., Mawritz, M. B., & Keating, D. J. (2019). Can you handle the pressure? The effect of performance pressure on stress appraisals, self-regulation, and behavior. Academy of Management Journal, 62(2), 531-552.	No	No	No	Yes	No
4	AMJ	2019	Shin, J., & Grant, A. M. (2019). Bored by Interest: How Intrinsic Motivation in One Task Can Reduce Performance on Other Tasks. Academy of Management Journal, 62(2), 415-436.	Yes	No	No	Yes	No
5	AMJ	2019	Lu, S., Bartol, K. M., Venkataramani, V., Zheng, X., & Liu, X. (2019). Pitching novel ideas to the boss: The interactive effects of employees' idea enactment and influence tactics on creativity assessment and implementation. Academy of Management Journal, 62(2), 579-606.	Yes	No	No	Yes	No
6	AMJ	2019	Shea, C. T., & Hawn, O. V. (2019). Microfoundations of corporate social responsibility and irresponsibility. Academy of Management Journal, 62(5), 1609-1642.	Yes	No	No	Yes	Yes
7	AMJ	2019	Porck, J. P., Matta, F. K., Hollenbeck, J. R., Oh, J. K., Lanaj, K., & Lee, S. M. (2019). Social identification in multiteam systems: The role of depletion and task complexity. Academy of Management Journal, 62(4), 1137-1162.	No	No	No	Yes	No
8	AMJ	2019	Sherf, E. N., Venkataramani, V., & Gajendran, R. S. (2019). Too busy to be fair? The effect of workload and rewards on managers' justice rule adherence. Academy of Management Journal, 62(2), 469-502.	No	No	No	Yes	No
9	AMJ	2019	Brands, R. A., & Mehra, A. (2019). Gender, brokerage, and performance: a construal approach. Academy of Management Journal, 62(1), 196-219.	Yes	No	No	Yes	Yes
10	AMJ	2019	Venus, M., Stam, D., & Van Knippenberg, D. (2019). Visions of change as visions of continuity. Academy of Management Journal, 62(3), 667-690.	Yes	No	No	Yes	Yes
11	AMJ	2019	Antino, M., Rico, R., & Thatcher, S. M. (2019). Structuring Reality Through the Faultlines Lens: The Effects of Structure, Fairness, and Status Conflict on the Activated Faultlines–Performance Relationship. Academy of Management Journal, 62(5), 1444-1470.	Yes	No	No	Yes	Yes
12	AMJ	2019	Kim, Y. J., & Toh, S. M. (2019). Stuck in the past? The influence of a leader's past cultural experience on group culture and positive and negative group deviance. Academy of Management Journal, 62(3), 944-969.	Yes	No	No	Yes	Yes
13	AMJ	2019	Krause, R., Wu, Z., Bruton, G. D., & Carter, S. M. (2019). The coercive isomorphism ripple effect: An investigation of nonprofit interlocks on corporate boards. Academy of Management Journal, 62(1), 283-308.	Yes	No	No	Yes	No
14	AMJ	2019	Livne-Ofer, E., Coyle-Shapiro, J. A., & Pearce, J. L. (2019). Eyes Wide Open: Perceived Exploitation and Its Consequences. Academy of Management Journal, 62(6), 1989-2018.	No	No	No	Yes	No
15	AMJ	2019	Han, J. H., Kang, S., Oh, I. S., Kehoe, R. R., & Lepak, D. P. (2019). The Goldilocks Effect of Strategic Human Resource Management? Optimizing the Benefits of a High-Performance Work System Through the Dual Alignment of Vertical and Horizontal Fit. Academy of Management Journal, 62(5), 1388-1412.	Yes	No	No	Yes	No

No	Journal	Year	Reference	OLS	Outliers	Model assumptions	Bootstrap	PROCESS macro
16	AMJ	2019	Ehrhardt, K., & Ragins, B. R. (2019). Relational attachment at work: A complementary fit perspective on the role of relationships in organizational life. Academy of Management Journal, 62(1), 248-282.	Yes	Yes	No	Yes	No
17	AMJ	2019	Hussain, I., Shu, R., Tangirala, S., & Ekkirala, S. (2019). The voice bystander effect: How information redundancy inhibits employee voice. Academy of Management Journal, 62(3), 828-849.	No	No	No	Yes	No
18	SMJ	2019	Huang, T. Y., Souitaris, V., & Barsade, S. G. (2019). Which matters more? Group fear versus hope in entrepreneurial escalation of commitment. Strategic Management Journal, 40(11), 1852-1881.	Yes	No	No	Yes	Yes
19	SMJ	2019	Garg, S., John Li, Q., & Shaw, J. D. (2019). Entrepreneurial firms grow up: Board undervaluation, board evolution, and firm performance in newly public firms. Strategic Management Journal, 40(11), 1882-1907.	No	No	No	No	No
20	SMJ	2019	Petrenko, O. V., Aime, F., Recendes, T., & Chandler, J. A. (2019). The case for humble expectations: CEO humility and market performance. Strategic Management Journal, 40(12), 1938-1964.	Yes	No	No	No	No
21	SMJ	2019	Wang, L., Wu, B., Pechmann, C., & Wang, Y. (2019). The performance effects of creative imitation on original products: Evidence from lab and field experiments. Strategic Management Journal.	Yes	No	No	Yes	No
22	SMJ	2019	Li, J., Li, P., & Wang, B. (2019). The liability of opaqueness: State ownership and the likelihood of deal completion in international acquisitions by Chinese firms. Strategic Management Journal, 40(2), 303-327.	No	No	No	Yes	No
23	SMJ	2019	Westphal, J. D., & Zhu, D. H. (2019). Under the radar: How firms manage competitive uncertainty by appointing friends of other chief executive officers to their boards. Strategic Management Journal, 40(1), 79-107.	No	No	No	Yes	No
24	SMJ	2019	Hill, A. D., Recendes, T., & Ridge, J. W. (2019). Second-order effects of CEO characteristics: How rivals' perceptions of CEOs as submissive and provocative precipitate competitive attacks. Strategic Management Journal, 40(5), 809-835.	No	No	No	No	No
25	JAP	2019	Ng, T. W., & Yam, K. C. (2019). When and why does employee creativity fuel deviance? Key psychological mechanisms. Journal of Applied Psychology, 104(9), 1144.	No	No	No	No	No
26	JAP	2019	Moore, C., Mayer, D. M., Chiang, F. F., Crossley, C., Karlesky, M. J., & Birtch, T. A. (2019). Leaders matter morally: The role of ethical leadership in shaping employee moral cognition and misconduct. Journal of Applied Psychology, 104(1), 123.	Yes	No	No	Yes	Yes
27	JAP	2019	Evans, J. B., Slaughter, J. E., Ellis, A. P., & Rivin, J. M. (2019). Gender and the evaluation of humor at work. Journal of Applied Psychology.	Yes	No	No	Yes	No
28	JAP	2019	Lievens, F., Sackett, P. R., Dahlke, J. A., Oostrom, J. K., & De Soete, B. (2019). Constructed response formats and their effects on minority–majority differences and validity. Journal of Applied Psychology, 104(5), 715.	Yes	No	No	Yes	Yes
29	JAP	2019	Wang, L., Law, K. S., Zhang, M. J., Li, Y. N., & Liang, Y. (2019). It’s mine! Psychological ownership of one’s job explains positive and negative workplace outcomes of job engagement. Journal of Applied Psychology, 104(2), 229.	No	No	No	Yes	No
30	JAP	2019	Bindl, U. K., Unsworth, K. L., Gibson, C. B., & Stride, C. B. (2019). Job crafting revisited: Implications of an extended framework for active changes at work. Journal of Applied Psychology, 104(5), 605.	Yes	No	No	Yes	No
31	JAP	2019	Rosen, C. C., Simon, L. S., Gajendran, R. S., Johnson, R. E., Lee, H. W., & Lin, S. H. J. (2019). Boxed in by your inbox: Implications of daily e-mail demands for managers’ leadership behaviors. Journal of Applied Psychology, 104(1), 19.	No	No	No	No	No
32	JAP	2019	Owens, B. P., Yam, K. C., Bednar, J. S., Mao, J., & Hart, D. W. (2019). The impact of leader moral humility on follower moral self-efficacy and behavior. Journal of Applied Psychology, 104(1), 146.	Yes	No	No	Yes	Yes

No	Journal	Year	Reference	OLS	Outliers	Model assumptions	Bootstrap	PROCESS macro
33	JAP	2019	Diefendorff, J. M., Gabriel, A. S., Nolan, M. T., & Yang, J. (2019). Emotion regulation in the context of customer mistreatment and felt affect: An event-based profile approach. Journal of Applied Psychology, 104(7), 965.	No	No	No	Yes	No
34	JAP	2019	Huang, Y. S. S., Greenbaum, R. L., Bonner, J. M., & Wang, C. S. (2019). Why sabotage customers who mistreat you? Activated hostility and subsequent devaluation of targets as a moral disengagement mechanism. Journal of Applied Psychology, 104(4), 495.	No	No	No	Yes	No
35	JAP	2019	Lee, H. W., Bradburn, J., Johnson, R. E., Lin, S. H. J., & Chang, C. H. D. (2019). The benefits of receiving gratitude for helpers: A daily investigation of proactive and reactive helping at work. Journal of Applied Psychology, 104(2), 197.	No	Yes	No	No	No
36	JAP	2019	Porter, C. M., Woo, S. E., Allen, D. G., & Keith, M. G. (2019). How do instrumental and expressive network positions relate to turnover? A meta-analytic investigation. Journal of Applied Psychology, 104(4), 511.	No	No	No	No	No
37	JAP	2019	Hernandez, M., Avery, D. R., Volpone, S. D., & Kaiser, C. R. (2019). Bargaining while Black: The role of race in salary negotiations. Journal of Applied Psychology, 104(4), 581.	Yes	No	No	No	No
38	JAP	2019	Koopmann, J., Johnson, R. E., Wang, M., Lanaj, K., Wang, G., & Shi, J. (2019). A self-regulation perspective on how and when regulatory focus differentially relates to citizenship behaviors. Journal of Applied Psychology, 104(5), 629.	No	No	No	Yes	No
39	JAP	2019	Taylor, S. G., Griffith, M. D., Vadera, A. K., Folger, R., & Letwin, C. R. (2019). Breaking the cycle of abusive supervision: How disidentification and moral identity help the trickle-down change course. Journal of Applied Psychology, 104(1), 164.	No	Yes	No	Yes	No
40	JAP	2019	Liao, Z., Liu, W., Li, X., & Song, Z. (2019). Give and take: An episodic perspective on leader-member exchange. Journal of Applied Psychology, 104(1), 34.	No	No	No	Yes	No
41	JAP	2019	Koopman, J., Scott, B. A., Matta, F. K., Conlon, D. E., & Dennerlein, T. (2019). Ethical leadership as a substitute for justice enactment: An information-processing perspective. Journal of Applied Psychology, 104(9), 1103.	No	No	No	Yes	No
42	JAP	2019	Lennard, A. C., Scott, B. A., & Johnson, R. E. (2019). Turning frowns (and smiles) upside down: A multilevel examination of surface acting positive and negative emotions on well-being. Journal of Applied Psychology, 104(9), 1164.	No	No	No	Yes	No
43	JAP	2019	Chen, G., Smith, T. A., Kirkman, B. L., Zhang, P., Lemoine, G. J., & Farh, J. L. (2019). Multiple team membership and empowerment spillover effects: Can empowerment processes cross team boundaries?. Journal of Applied Psychology, 104(3), 321.	No	Yes	No	Yes	No
44	JAP	2019	Carlson, D. S., Thompson, M. J., & Kacmar, K. M. (2019). Double crossed: The spillover and crossover effects of work demands on work outcomes through the family. Journal of Applied Psychology, 104(2), 214.	No	Yes	Yes	No	No
45	JAP	2019	Mayer, D. M., Ong, M., Sonenshein, S., & Ashford, S. J. (2019). The money or the morals? When moral language is more effective for selling social issues. Journal of Applied Psychology, 104(8), 1058.	Yes	Yes	No	Yes	Yes
46	JAP	2019	Gündemir, S., Carton, A. M., & Homan, A. C. (2019). The impact of organizational performance on the emergence of Asian American leaders. Journal of Applied Psychology, 104(1), 107.	Yes	No	No	Yes	Yes
47	JAP	2019	Sitzmann, T., Ployhart, R. E., & Kim, Y. (2019). A process model linking occupational strength to attitudes and behaviors: The explanatory role of occupational personality heterogeneity. Journal of Applied Psychology, 104(2), 247.	No	No	No	Yes	No
48	JAP	2019	Liu, Z., Riggio, R. E., Day, D. V., Zheng, C., Dai, S., & Bian, Y. (2019). Leader development begins at home: Overparenting harms adolescent leader emergence. Journal of Applied Psychology, 104(10), 1226.	No	Yes	No	Yes	No
49	JAP	2019	Zhou, Y., Zou, M., Woods, S. A., & Wu, C. H. (2019). The restorative effect of work after unemployment: An intraindividual analysis of subjective well-being recovery through reemployment. Journal of Applied Psychology, 104(9), 1195.	Yes	No	No	Yes	No

No	Journal	Year	Reference	OLS	Outliers	Model assumptions	Bootstrap	PROCESS macro
50	JAP	2019	Booth-LeDoux, S. M., Matthews, R. A., & Wayne, J. H. (2019). Testing a resource-based spillover-crossover-spillover model: Transmission of social support in dual-earner couples. Journal of Applied Psychology.	No	No	No	Yes	No
51	JAP	2019	Lanaj, K., Foulk, T. A., & Erez, A. (2019). Energizing leaders via self-reflection: A within-person field experiment. Journal of Applied Psychology, 104(1), 1.	No	No	No	Yes	No
52	JAP	2019	Hulshof, I. L., Demerouti, E., & Le Blanc, P. M. (2019). Reemployment crafting: Proactively shaping one's job search. Journal of Applied Psychology.	No	No	No	Yes	No
53	JAP	2019	Sessions, H., Nahrgang, J. D., Newton, D. W., & Chamberlin, M. (2019). I'm tired of listening: The effects of supervisor appraisals of group voice on supervisor emotional exhaustion and performance. Journal of Applied Psychology.	Yes	No	No	Yes	Yes
54	JAP	2019	Lin, K. J., Savani, K., & Ilies, R. (2019). Doing good, feeling good? The roles of helping motivation and citizenship pressure. Journal of Applied Psychology.	Yes	No	No	Yes	Yes
55	JAP	2019	Qin, X., Chen, C., Yam, K. C., Huang, M., & Ju, D. (2019). The double-edged sword of leader humility: Investigating when and why leader humility promotes versus inhibits subordinate deviance. Journal of Applied Psychology.	Yes	No	No	Yes	No
56	JAP	2019	Priesemuth, M., & Bigelow, B. (2019). It hurts me too!(or not?): Exploring the negative implications for abusive bosses. Journal of Applied Psychology.	Yes	No	No	Yes	No
57	JAP	2019	Cowen, A. P., & Montgomery, N. V. (2019). To be or not to be sorry? How CEO gender impacts the effectiveness of organizational apologies. Journal of Applied Psychology.	Yes	No	No	Yes	Yes
58	JAP	2019	McCarthy, J. E., & Levin, D. Z. (2019). Network residues: The enduring impact of intra-organizational dormant ties. Journal of Applied Psychology.	Yes	Yes	No	Yes	No
59	JAP	2019	Hall, E. V., Avery, D. R., McKay, P. F., Blot, J. F., & Edwards, M. (2019). Composition and compensation: The moderating effect of individual and team performance on the relationship between Black team member representation and salary. Journal of Applied Psychology, 104(3), 448.	No	No	No	Yes	No
60	JAP	2019	Burmeister, A., Wang, M., & Hirschi, A. (2019). Understanding the motivational benefits of knowledge transfer for older and younger workers in age-diverse coworker dyads: An actor–partner interdependence model. Journal of applied psychology.	No	No	No	Yes	No
61	JAP	2019	Maltarich, M. A., Reilly, G., & DeRose, C. (2019). A theoretical assessment of dismissal rates and unit performance, with empirical evidence. Journal of Applied Psychology.	No	No	No	Yes	No
62	JAP	2019	Rapp, T. L., & Mathieu, J. E. (2019). Team and individual influences on members' identification and performance per membership in multiple team membership arrangements. Journal of Applied Psychology, 104(3), 303.	No	No	No	Yes	No
63	JAP	2019	Parker, S. K., Andrei, D. M., & Van den Broeck, A. (2019). Poor work design begets poor work design: Capacity and willingness antecedents of individual work design behavior. Journal of Applied Psychology.	Yes	Yes	No	Yes	Yes
64	JAP	2019	Mohr, J. J., Markell, H. M., King, E. B., Jones, K. P., Peddie, C. I., & Kendra, M. S. (2019). Affective antecedents and consequences of revealing and concealing a lesbian, gay, or bisexual identity. Journal of Applied Psychology.	No	No	Yes	Yes	No
65	JAP	2019	Yu, K. Y. T. (2019). Influencing how one is seen by potential talent: Organizational impression management among recruiting firms. Journal of Applied Psychology.	Yes	No	No	Yes	Yes
66	JAP	2019	Hu, J., Zhang, Z., Jiang, K., & Chen, W. (2019). Getting ahead, getting along, and getting prosocial: Examining extraversion facets, peer reactions, and leadership emergence. Journal of Applied Psychology.	No	No	No	Yes	No

No	Journal	Year	Reference	OLS	Outliers	Model assumptions	Bootstrap	PROCESS macro
67	JAP	2019	Ouyang, K., Cheng, B. H., Lam, W., & Parker, S. K. (2019). Enjoy your evening, be proactive tomorrow: How off-job experiences shape daily proactivity. Journal of Applied Psychology, 104(8), 1003.	No	No	No	Yes	No
68	JAP	2019	Zhang, Y., Zhang, Y., Ng, T. W., & Lam, S. S. (2019). Promotion-and prevention-focused coping: A meta-analytic examination of regulatory strategies in the work stress process. Journal of Applied Psychology, 104(10), 1296.	No	No	No	Yes	No
69	JAP	2019	Jacob, G. H., Frese, M., Krauss, S. I., & Friedrich, C. (2019). On the importance of a motivational agency variable: Being a formal business in developing countries is only helpful for growth if business owners show a high degree of personal initiative. Journal of Applied Psychology, 104(9), 1181.	Yes	Yes	No	Yes	Yes
70	JAP	2019	Foulk, T. A., Lanaj, K., & Krishnan, S. (2019). The virtuous cycle of daily motivation: Effects of daily strivings on work behaviors, need satisfaction, and next-day strivings. Journal of Applied Psychology, 104(6), 755.	No	No	No	Yes	No
71	JAP	2019	Beck, J. W., Schmidt, A. M., & Natali, M. W. (2019). Efficient proximal resource allocation strategies predict distal team performance: Evidence from the National Hockey League. Journal of Applied Psychology.	No	No	No	Yes	No
72	Org. Sci.	2019	Reiche, B. S., & Neeley, T. B. (2019). Head, Heart, or Hands: How Do Employees Respond to a Radical Global Language Change over Time?. Organization Science, 30(6), 1252-1269.	No	No	No	Yes	No
73	Org. Sci.	2019	Chatman, J. A., Greer, L. L., Sherman, E., & Doerr, B. (2019). Blurred lines: How the collectivism norm operates through perceived group diversity to boost or harm group performance in Himalayan mountain climbing. Organization Science, 30(2), 235-259.	Yes	No	No	Yes	Yes
74	Org. Sci.	2019	DesJardine, M., & Bansal, P. (2019). One step forward, two steps back: How negative external evaluations can shorten organizational time horizons. Organization Science, 30(4), 761-780.	No	No	No	No	No
75	Org. Sci.	2019	Radoynovska, N., & King, B. G. (2019). To Whom Are You True? Audience Perceptions of Authenticity in Nascent Crowdfunding Ventures. Organization Science, 30(4), 781-802.	Yes	No	No	Yes	Yes
76	Org. Sci.	2019	Baker, B., Derfler-Rozin, R., Pitesa, M., & Johnson, M. (2019). Stock Market Responses to Unethical Behavior in Organizations: An Organizational Context Model. Organization Science, 30(2), 319-336.	No	No	No	Yes	No
77	Org. Sci.	2019	Carson Marr, J., Pettit, N., & Thau, S. (2019). After the Fall: How Perceived Self-Control Protects the Legitimacy of Higher-Ranking Employees after Status Loss. Organization Science, 30(6), 1165-1188.	Yes	No	No	Yes	Yes
78	Org. Sci.	2019	Sherf, E. N., Tangirala, S., & Venkataramani, V. (2019). Why Managers Do Not Seek Voice from Employees: The Importance of Managers' Personal Control and Long-Term Orientation. Organization Science, 30(3), 447-466.	Yes	Yes	No	Yes	Yes
79	Org. Sci.	2019	Wilhelm, H., Richter, A. W., & Semrau, T. (2019). Employee Learning from Failure: A Team-as-Resource Perspective. Organization Science, 30(4), 694-714.	No	No	No	Yes	No
80	Org. Sci.	2019	Lee, S. (2019). Learning-by-moving: can reconfiguring spatial proximity between organizational members promote individual-level exploration?. Organization Science, 30(3), 467-488.	Yes	No	No	No	No
81	Org. Sci.	2019	Moore, C. B., Payne, G. T., Filatotchev, I., & Zajac, E. J. (2019). The Cost of Status: When Social and Economic Interests Collide. Organization Science, 30(5), 869-884.	No	No	No	No	No
82	Org. Sci.	2019	Anderson, T., & Bidwell, M. (2019). Outside insiders: Understanding the role of contracting in the careers of managerial workers. Organization Science, 30(5), 1000-1029.	Yes	No	No	Yes	No
83	ASQ	2019	Galperin, R. V., Hahl, O., Sterling, A. D., & Guo, J. (2019). Too good to hire? Capability and inferences about commitment in labor markets. Administrative Science Quarterly, 0001839219840022.	Yes	No	No	Yes	Yes

No	Journal	Year	Reference	OLS	Outliers	Model assumptions	Bootstrap	PROCESS macro
84	ASQ	2019	DeCelles, K. A., Sonenshein, S., & King, B. G. (2019). Examining Anger's Immobilizing Effect on Institutional Insiders' Action Intentions in Social Movements. Administrative Science Quarterly, 0001839219879646.	Yes	No	No	Yes	Yes
85	ASQ	2019	Uribe, J., Sytch, M., & Kim, Y. H. (2019). When Friends Become Foes: Collaboration as a Catalyst for Conflict. Administrative Science Quarterly, 0001839219877507.	Yes	No	No	No	No
86	ASQ	2019	Inoue, C. (2019). Election Cycles and Organizations: How Politics Shapes the Performance of State-owned Enterprises over Time. Administrative Science Quarterly, 0001839219869913.	Yes	No	Yes	No	No

Figures

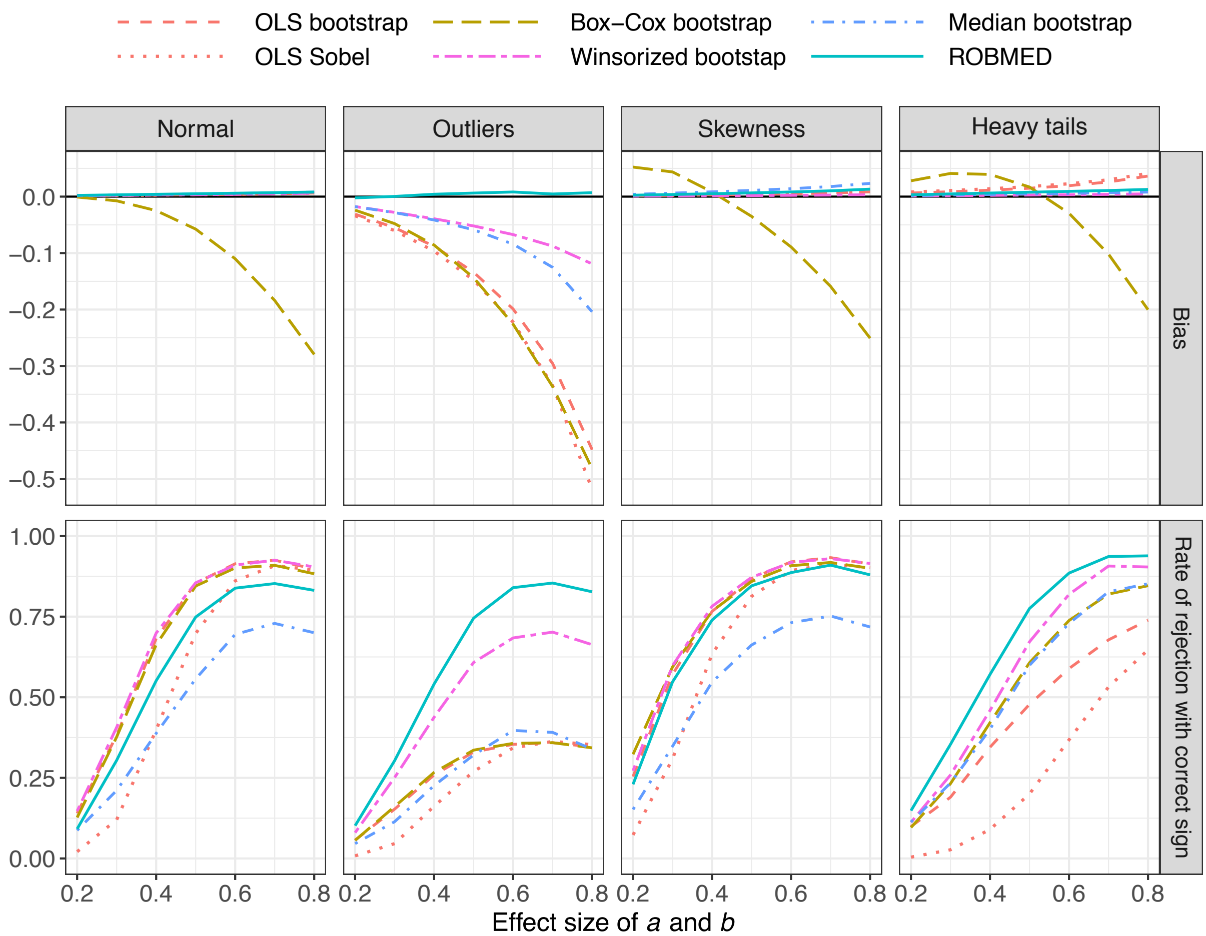

Figure 1. Results from 1000 simulation runs for simulation design 1 with varying effect sizes in the setting with mediation ($a = b = c = 0.2, \dots, 0.8$), sample size $n = 50$. The top row shows the average bias of the indirect effect and includes a horizontal reference line at 0 for no bias. The bottom row displays the rate of how often the methods reject the null hypothesis and the corresponding estimate of *ab* has the correct sign (a measure of realized power of the tests in the presence of outliers; the higher this rate the better).

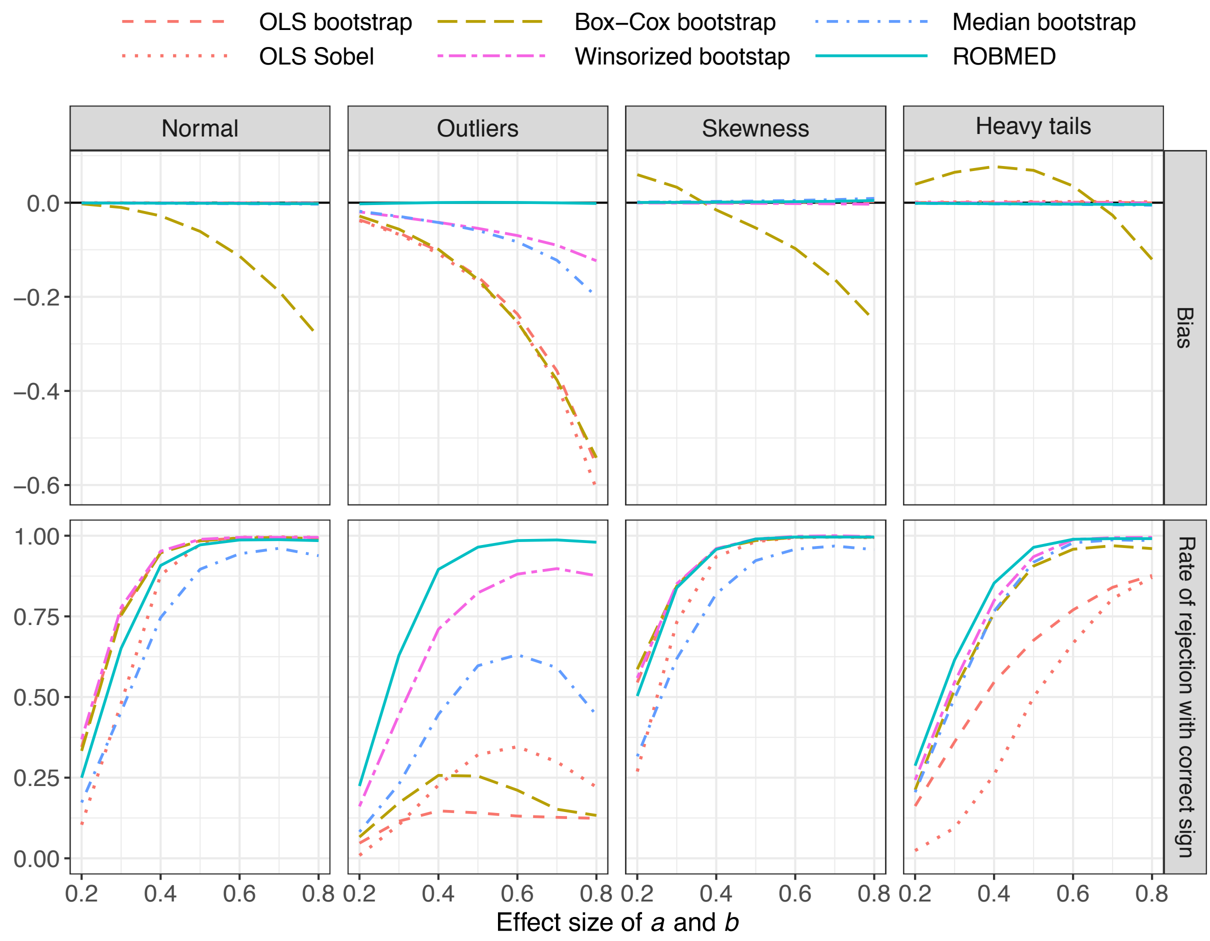

Figure 2. Results from 1000 simulation runs for simulation design 1 with varying effect sizes in the setting with mediation ($a = b = c = 0.2, \dots, 0.8$), and sample size $n = 100$. The top row shows the average bias of the indirect effect and includes a horizontal reference line at 0 for no bias. The bottom row displays the rate of how often the methods reject the null hypothesis and the corresponding estimate of *ab* has the correct sign (a measure of realized power of the tests in the presence of outliers; the higher this rate the better).

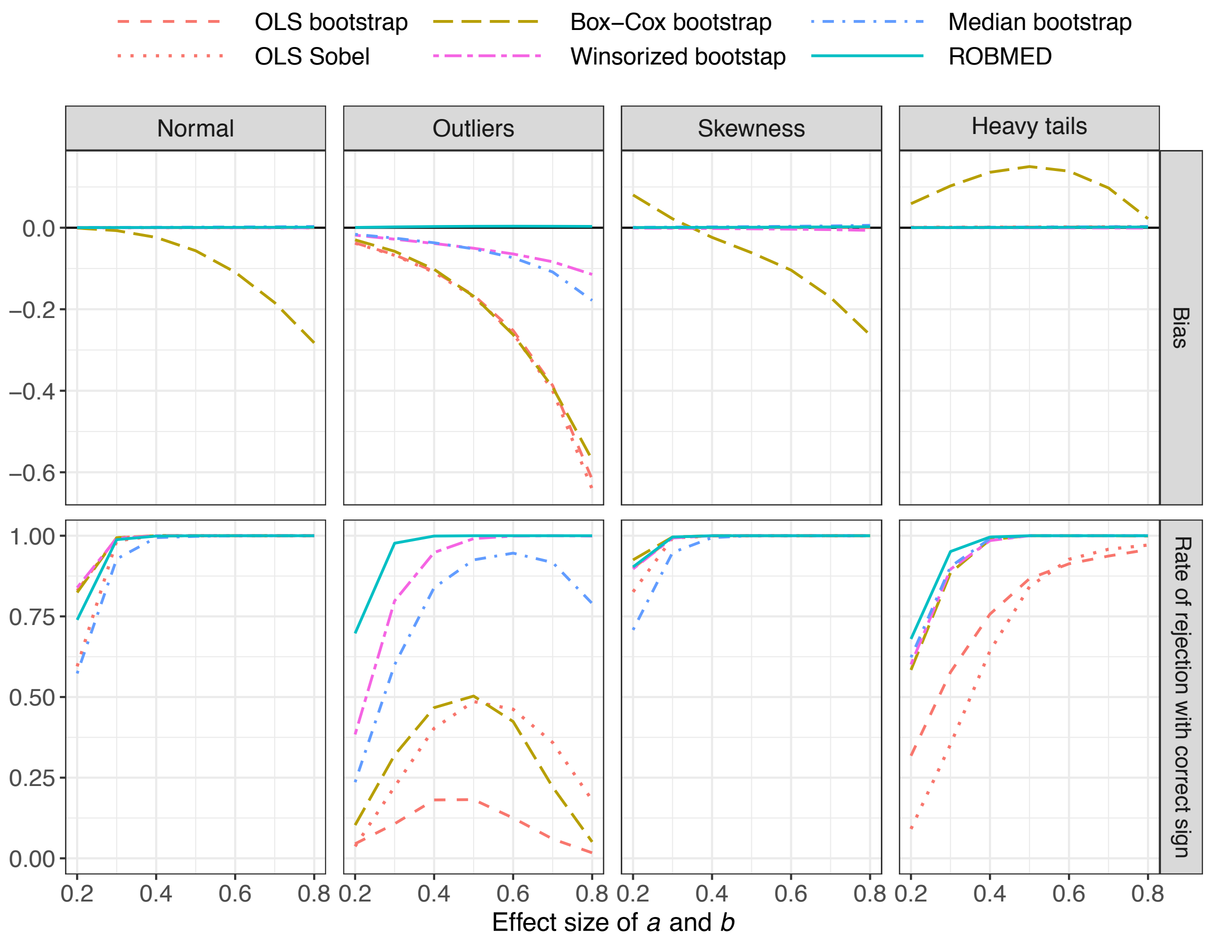

Figure 3. Results from 1000 simulation runs for simulation design 1 with varying effect sizes in the setting with mediation ($a = b = c = 0.2, \ldots, 0.8$), and sample size $n = 250$. The top row shows the average bias of the indirect effect and includes a horizontal reference line at 0 for no bias. The bottom row displays the rate of how often the methods reject the null hypothesis and the corresponding estimate of *ab* has the correct sign (a measure of realized power of the tests in the presence of outliers; the higher this rate the better).

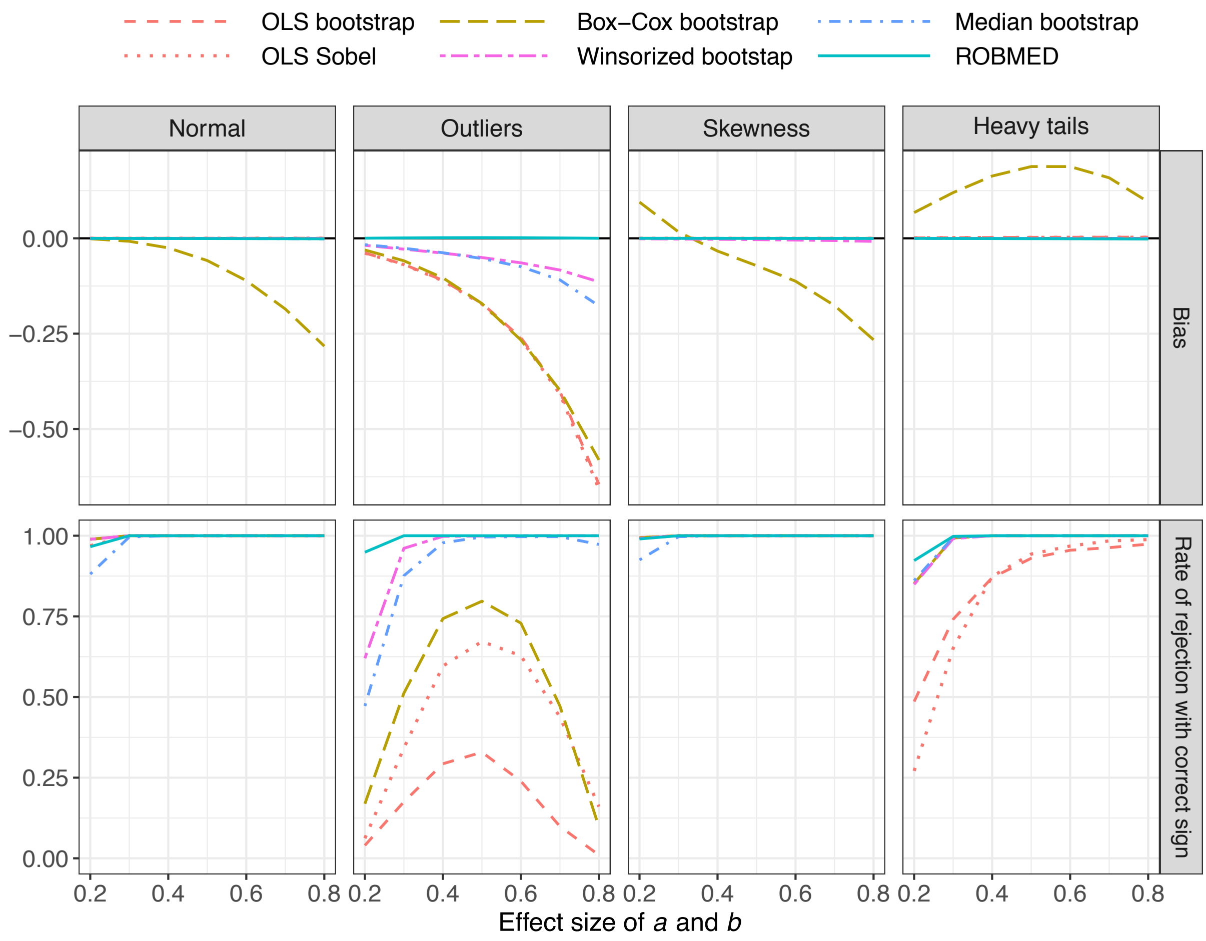

Figure 4. Results from 1000 simulation runs for simulation design 1 with varying effect sizes in the setting with mediation ($a = b = c = 0.2, \ldots, 0.8$), and sample size $n = 500$. The top row shows the average bias of the indirect effect and includes a horizontal reference line at 0 for no bias. The bottom row displays the rate of how often the methods reject the null hypothesis and the corresponding estimate of *ab* has the correct sign (a measure of realized power of the tests in the presence of outliers; the higher this rate the better).

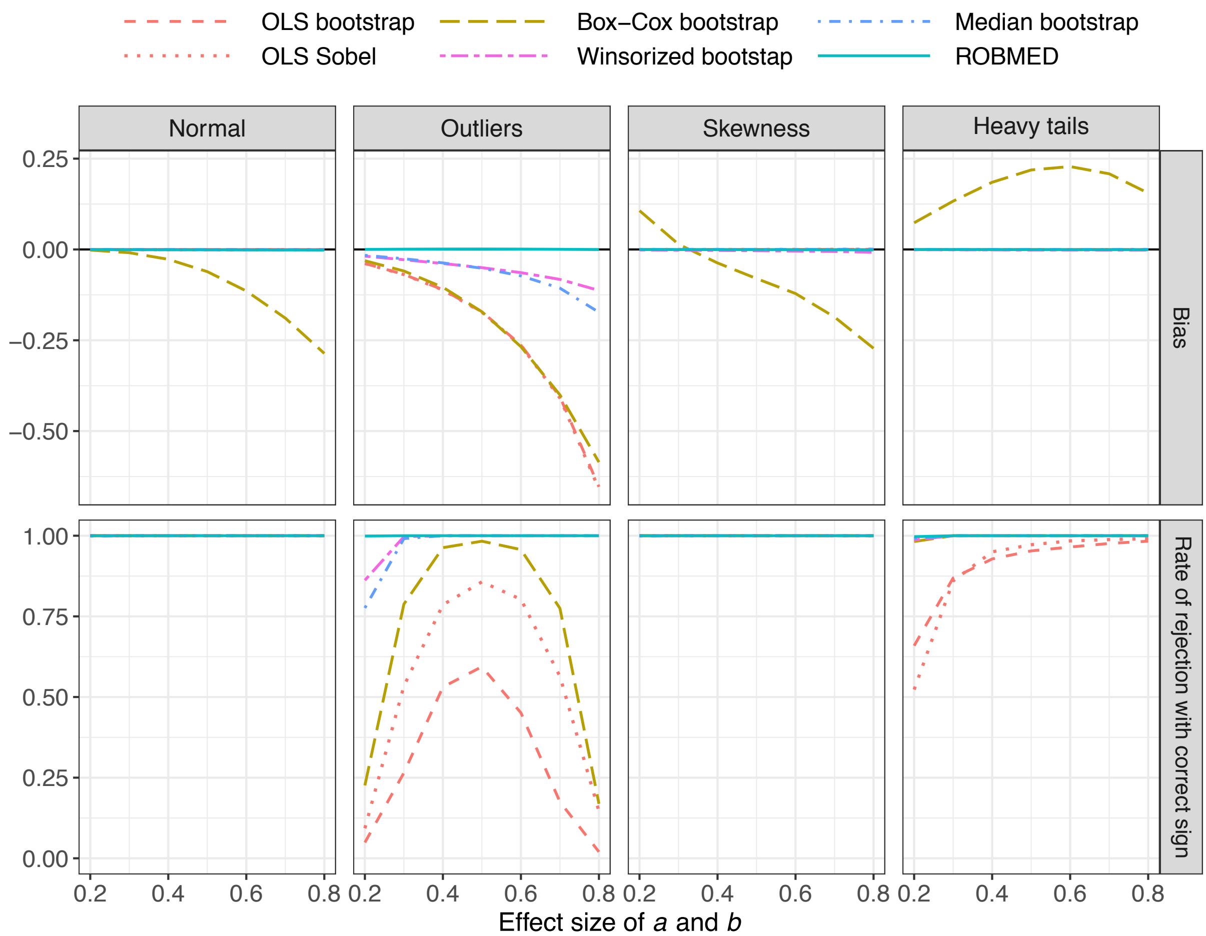

Figure 5. Results from 1000 simulation runs for simulation design 1 with varying effect sizes in the setting with mediation ($a = b = c = 0.2, \dots, 0.8$), and sample size $n = 1000$. The top row shows the average bias of the indirect effect and includes a horizontal reference line at 0 for no bias. The bottom row displays the rate of how often the methods reject the null hypothesis and the corresponding estimate of *ab* has the correct sign (a measure of realized power of the tests in the presence of outliers; the higher this rate the better).

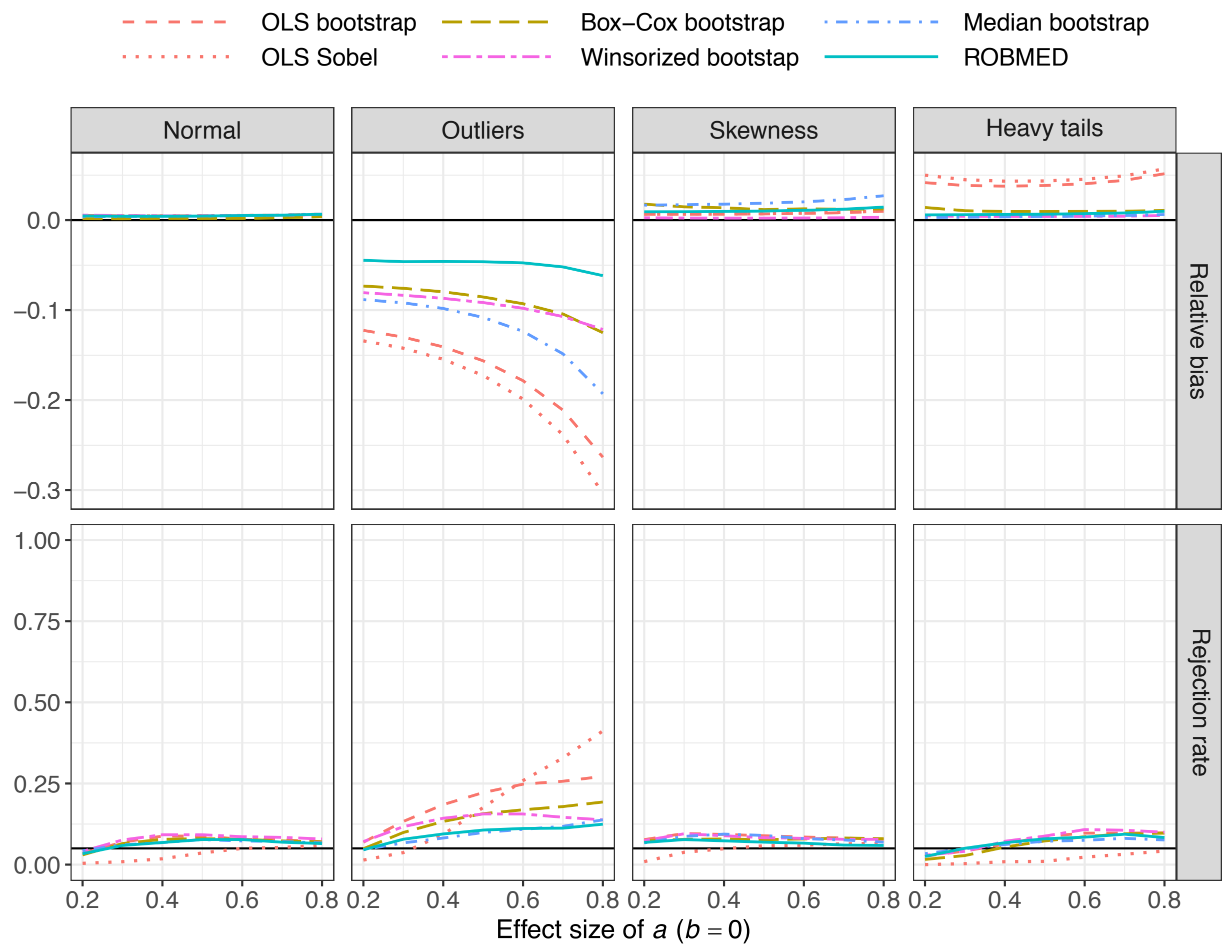

Figure 6. Results from 1000 simulation runs for simulation design 1 with varying effect sizes in the setting with no mediation ($a = c = 0.2, \dots, 0.8$, $b = 0$), and sample size $n = 50$. The top row shows the average relative bias of the indirect effect with respect to the effect size of a, and includes a horizontal reference line at 0 for no bias. The bottom row displays the rejection rate of the corresponding tests (i.e., the realized size), and a horizontal line is drawn for the nominal size $\alpha = 0.05$ (the closer to this line the better).

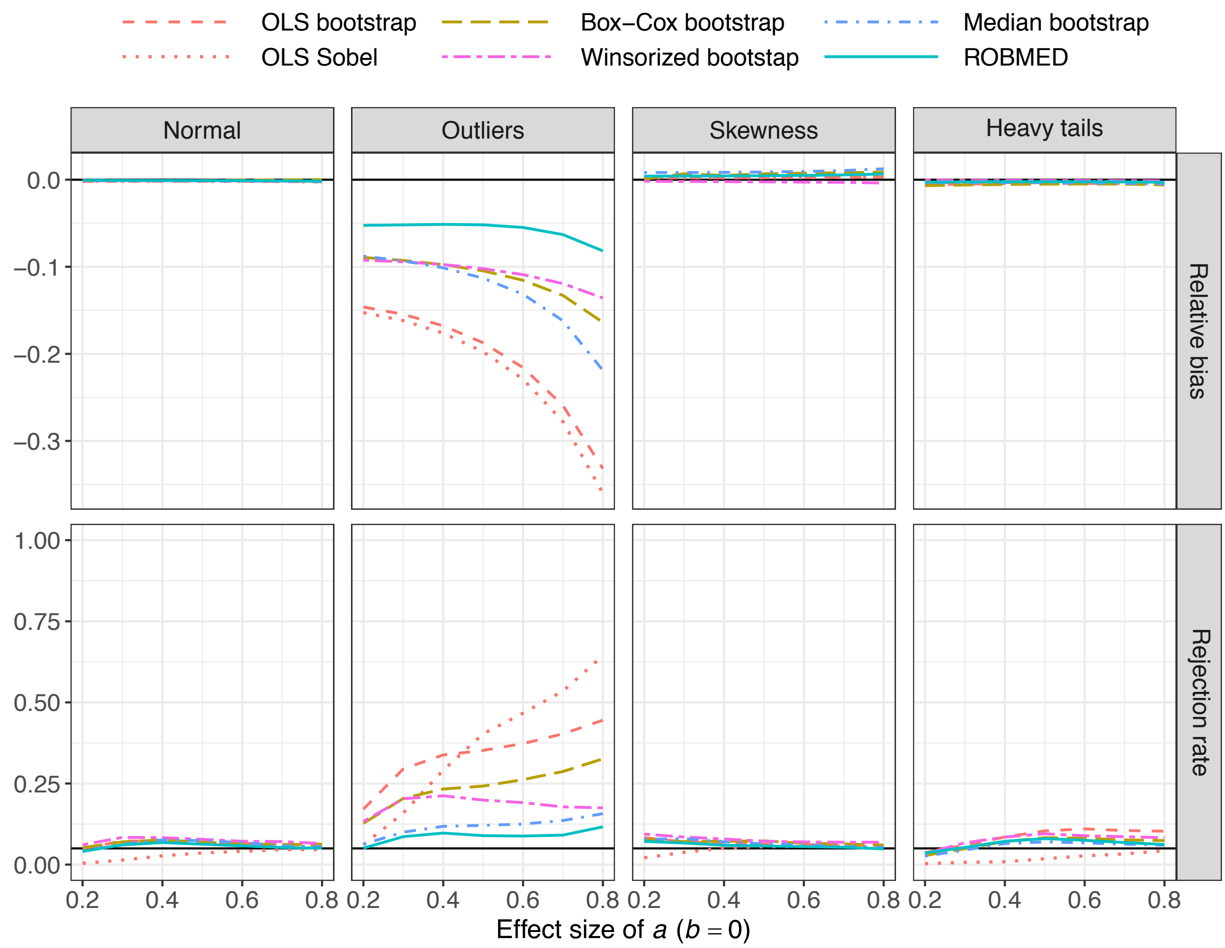

Figure 7. Results from 1000 simulation runs for simulation design 1 with varying effect sizes in the setting with no mediation ($a = c = 0.2, \dots, 0.8$, $b = 0$), and sample size $n = 100$. The top row shows the average relative bias of the indirect effect with respect to the effect size of a, and includes a horizontal reference line at 0 for no bias. The bottom row displays the rejection rate of the corresponding tests (i.e., the realized size), and a horizontal line is drawn for the nominal size $\alpha = 0.05$ (the closer to this line the better).

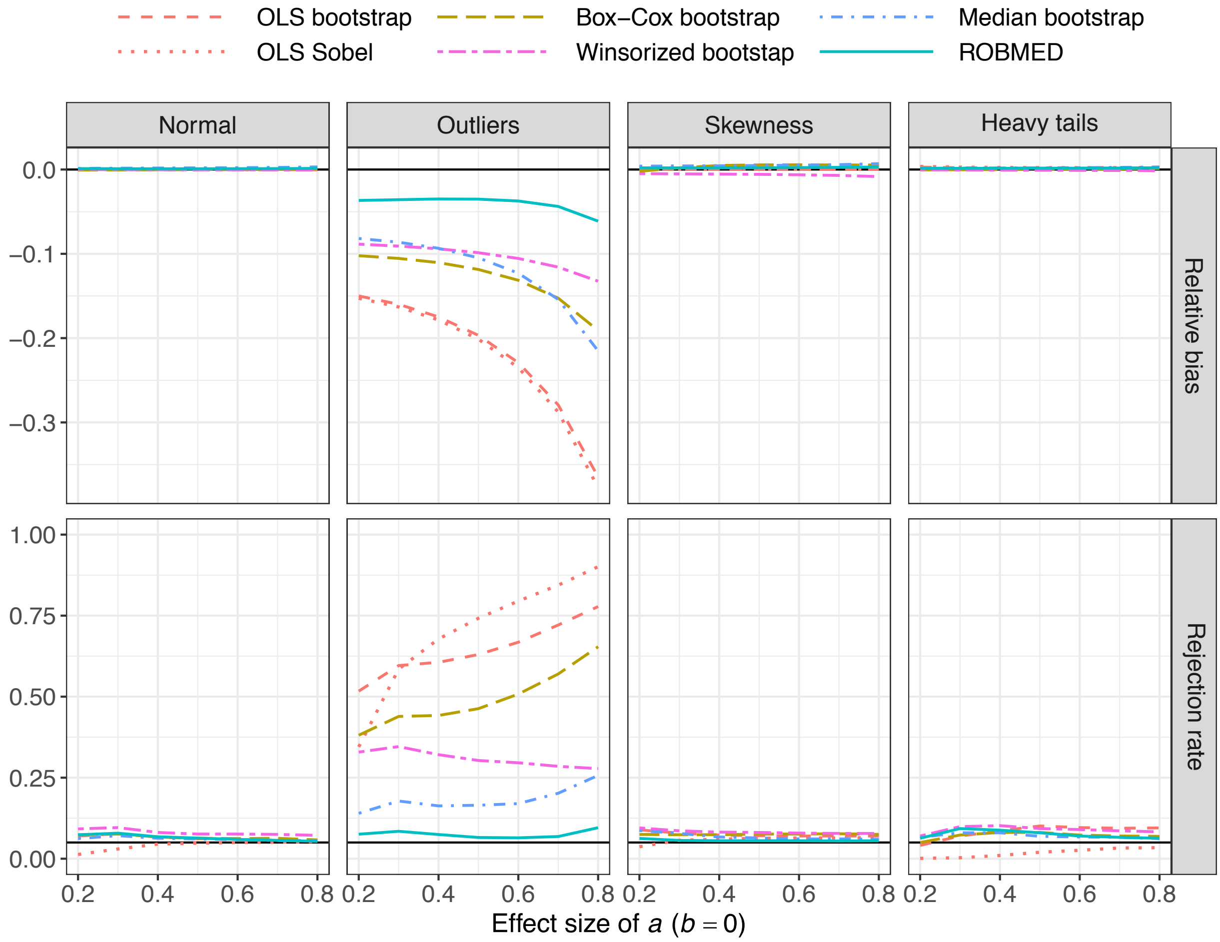

Figure 8. Results from 1000 simulation runs for simulation design 1 with varying effect sizes in the setting with no mediation ($a = c = 0.2, \dots, 0.8$, $b = 0$), and sample size $n = 250$. The top row shows the average relative bias of the indirect effect with respect to the effect size of a, and includes a horizontal reference line at 0 for no bias. The bottom row displays the rejection rate of the corresponding tests (i.e., the realized size), and a horizontal line is drawn for the nominal size $\alpha = 0.05$ (the closer to this line the better).

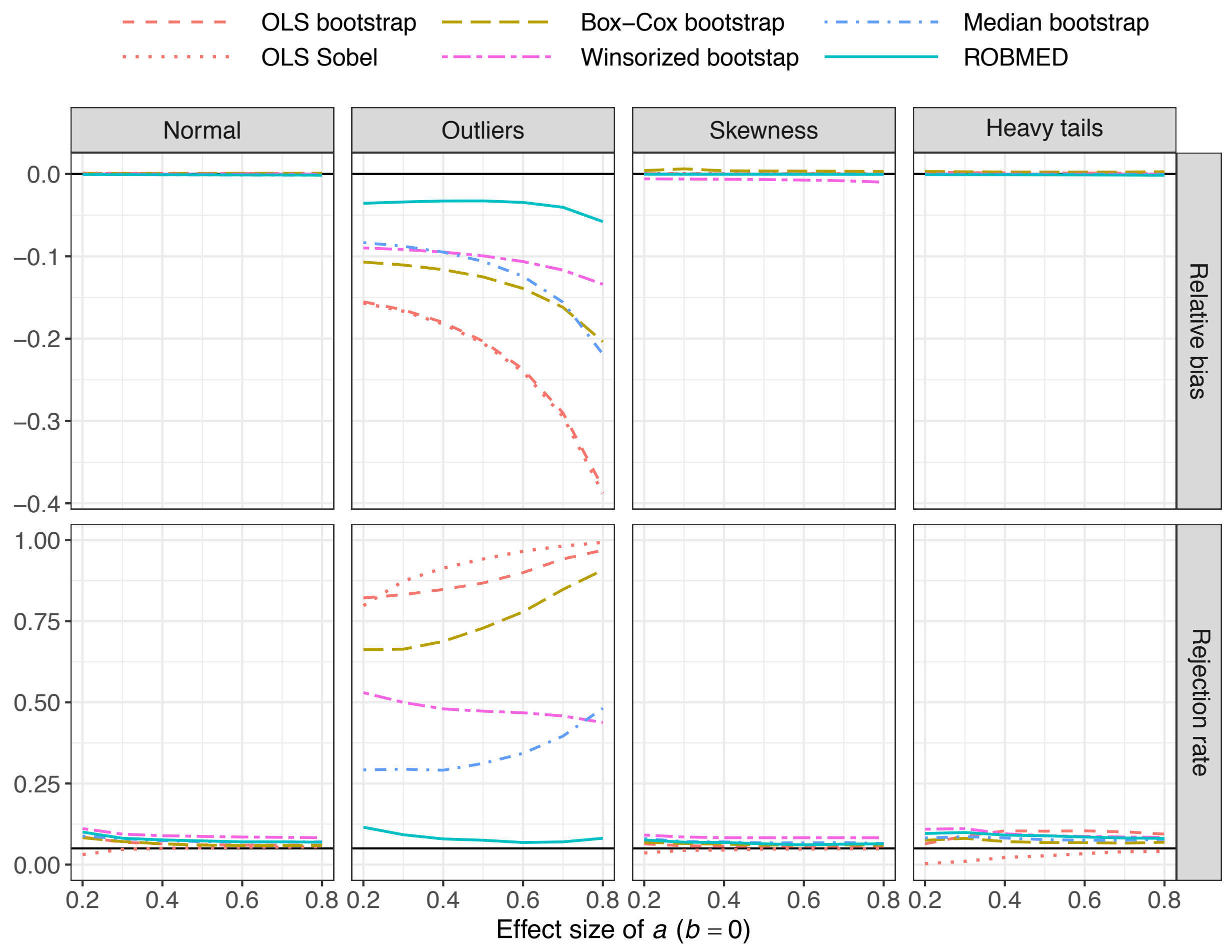

Figure 9. Results from 1000 simulation runs for simulation design 1 with varying effect sizes in the setting with no mediation ($a = c = 0.2, \ldots, 0.8$, $b = 0$), and sample size $n = 500$. The top row shows the average relative bias of the indirect effect with respect to the effect size of a, and includes a horizontal reference line at 0 for no bias. The bottom row displays the rejection rate of the corresponding tests (i.e., the realized size), and a horizontal line is drawn for the nominal size $\alpha = 0.05$ (the closer to this line the better).

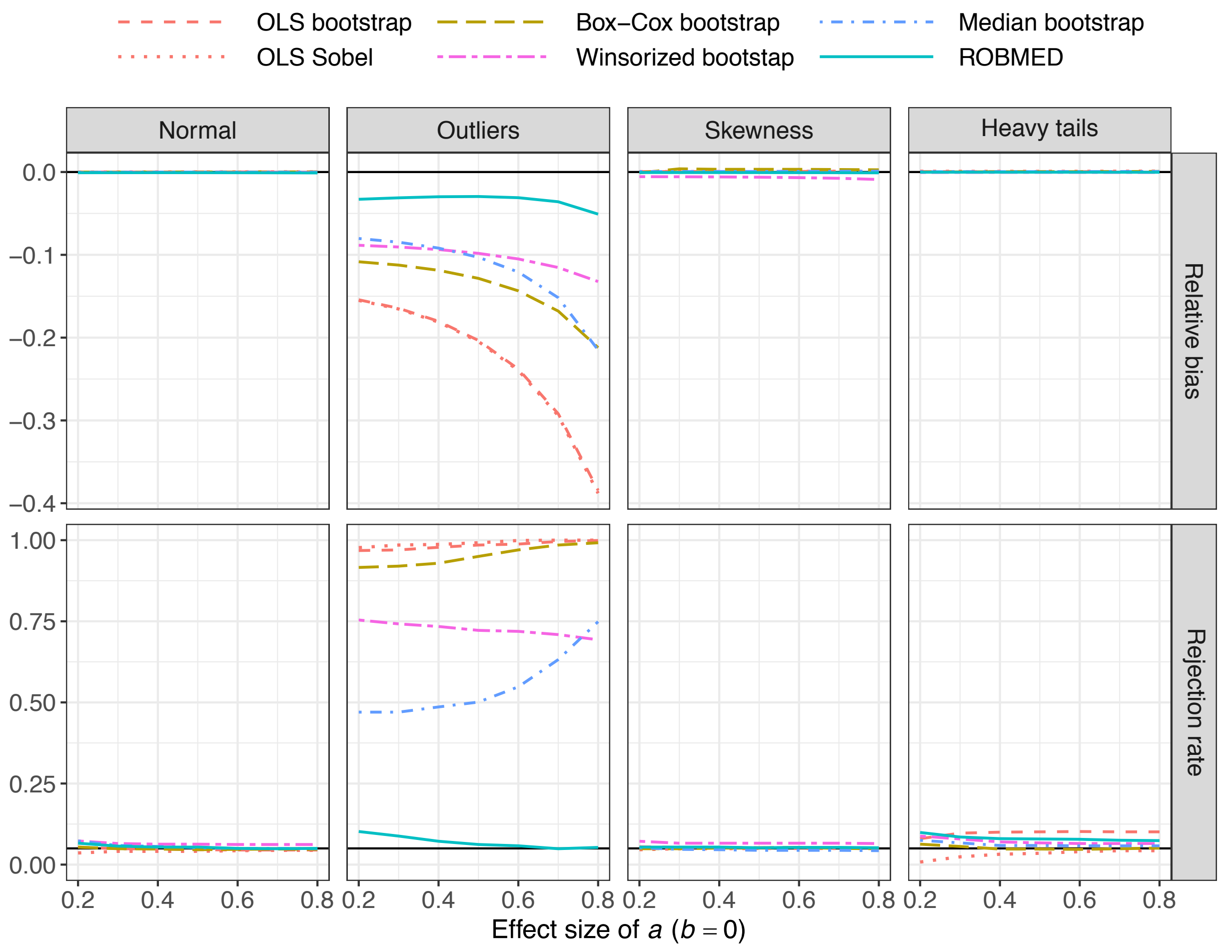

Figure 10. Results from 1000 simulation runs for simulation design 1 with varying effect sizes in the setting with no mediation ($a = c = 0.2, \dots, 0.8$, $b = 0$), and sample size $n = 1000$. The top row shows the average relative bias of the indirect effect with respect to the effect size of a, and includes a horizontal reference line at 0 for no bias. The bottom row displays the rejection rate of the corresponding tests (i.e., the realized size), and a horizontal line is drawn for the nominal size $\alpha = 0.05$ (the closer to this line the better).

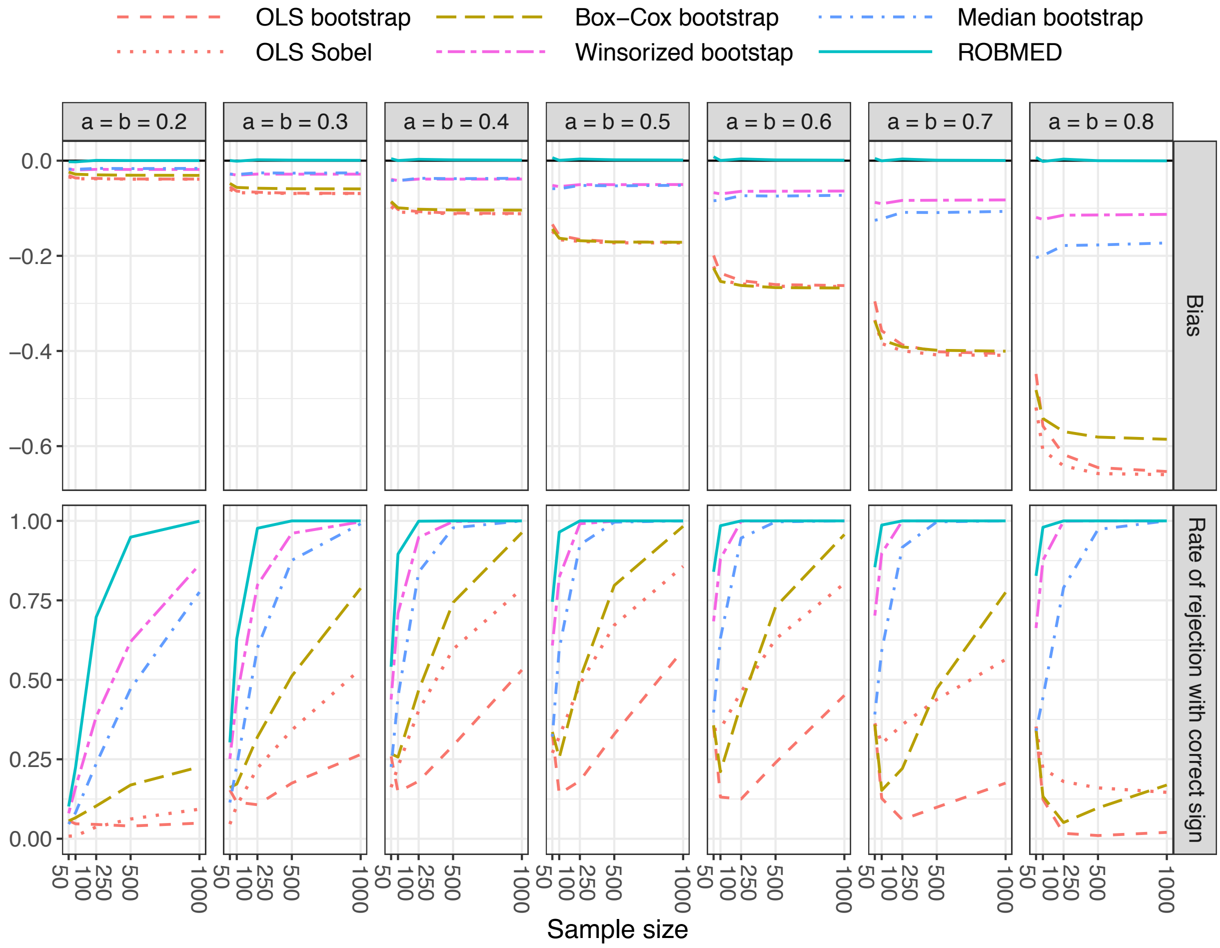

Figure 11. Results from 1000 simulation runs for simulation design 1 for the setting with outliers, varying sample size $n = 50, 100, 250, 500, 1000$, and varying effect sizes in the setting with mediation ($a = b = c = 0.2, \dots, 0.8$). The top row shows the average bias of the indirect effect and includes a horizontal reference line at 0 for no bias. The bottom row displays the rate of how often the methods reject the null hypothesis and the corresponding estimate of *ab* has the correct sign (a measure of realized power of the tests in the presence of outliers; the higher this rate the better).

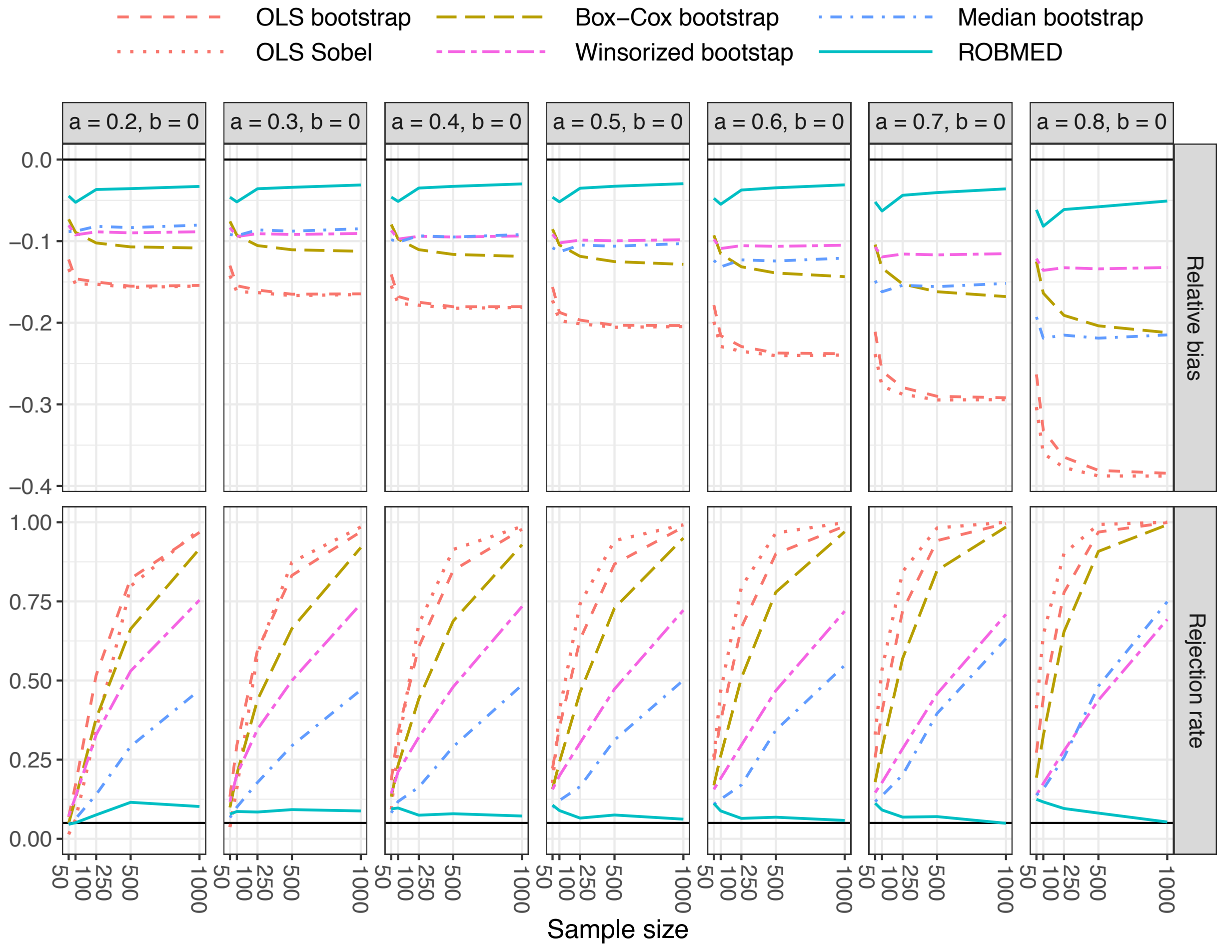

Figure 12. Results from 1000 simulation runs for simulation design 1 for the setting with outliers, varying sample size $n = 50, 100, 250, 500, 1000$, and varying effect sizes in the setting with no mediation ($a = c = 0.2, \dots, 0.8$, $b = 0$). The top row shows the average relative bias of the indirect effect with respect to the effect size of a, and includes a horizontal reference line at 0 for no bias. The bottom row displays the rejection rate of the corresponding tests (i.e., the realized size), and a horizontal line is drawn for the nominal size $\alpha = 0.05$ (the closer to this line the better).

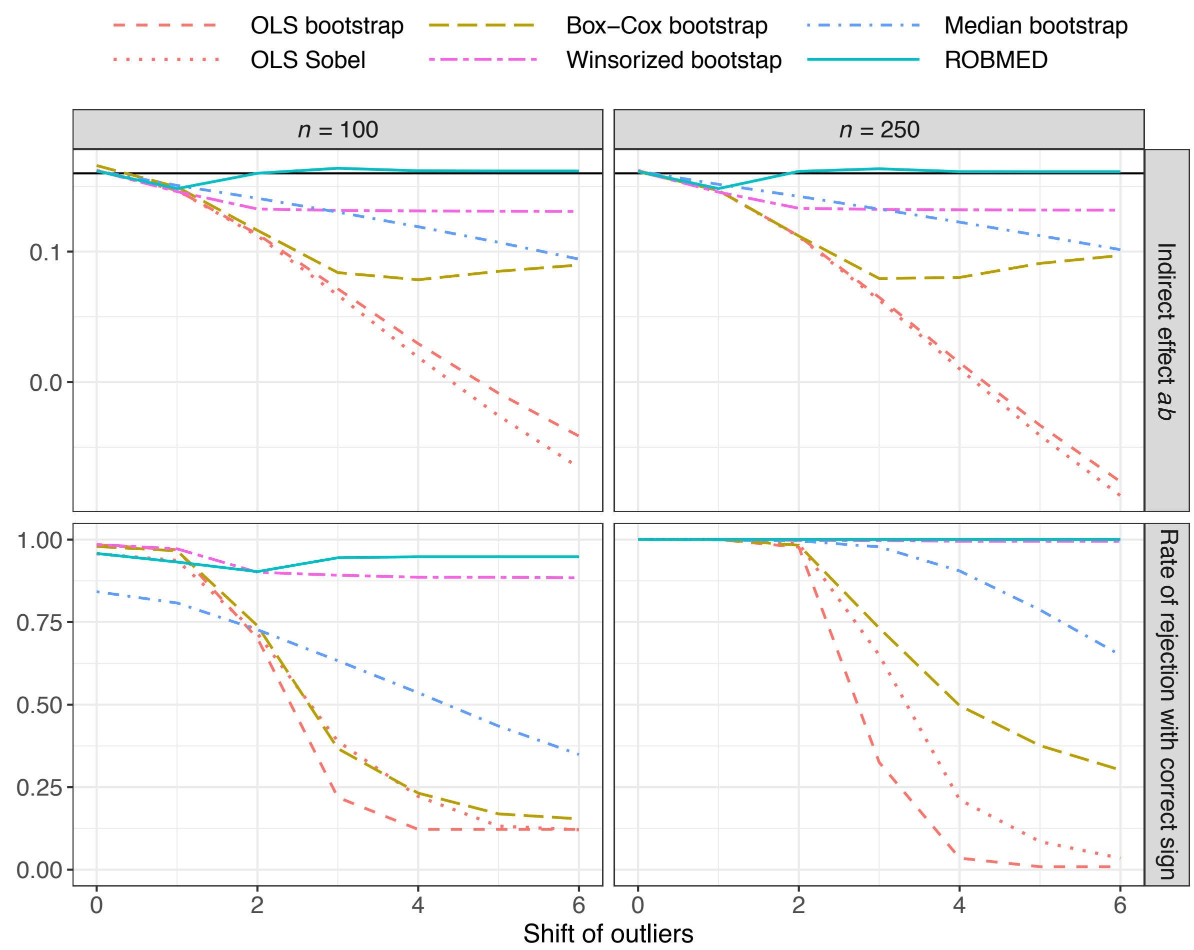

Figure 13. Results from 1000 simulation runs for simulation design 2 with varying distance of outliers (outlier probability 2%) and the setting with mediation ($a = 0.4$, $b = 0.4$). The top row shows the average estimates of the indirect effect and includes a horizontal reference line for the true indirect effect $ab = 0.16$. The bottom row displays the rate of how often the methods reject the null hypothesis and the corresponding estimate of *ab* has the correct sign (a measure of realized power of the tests in the presence of outliers; the higher this rate the better).

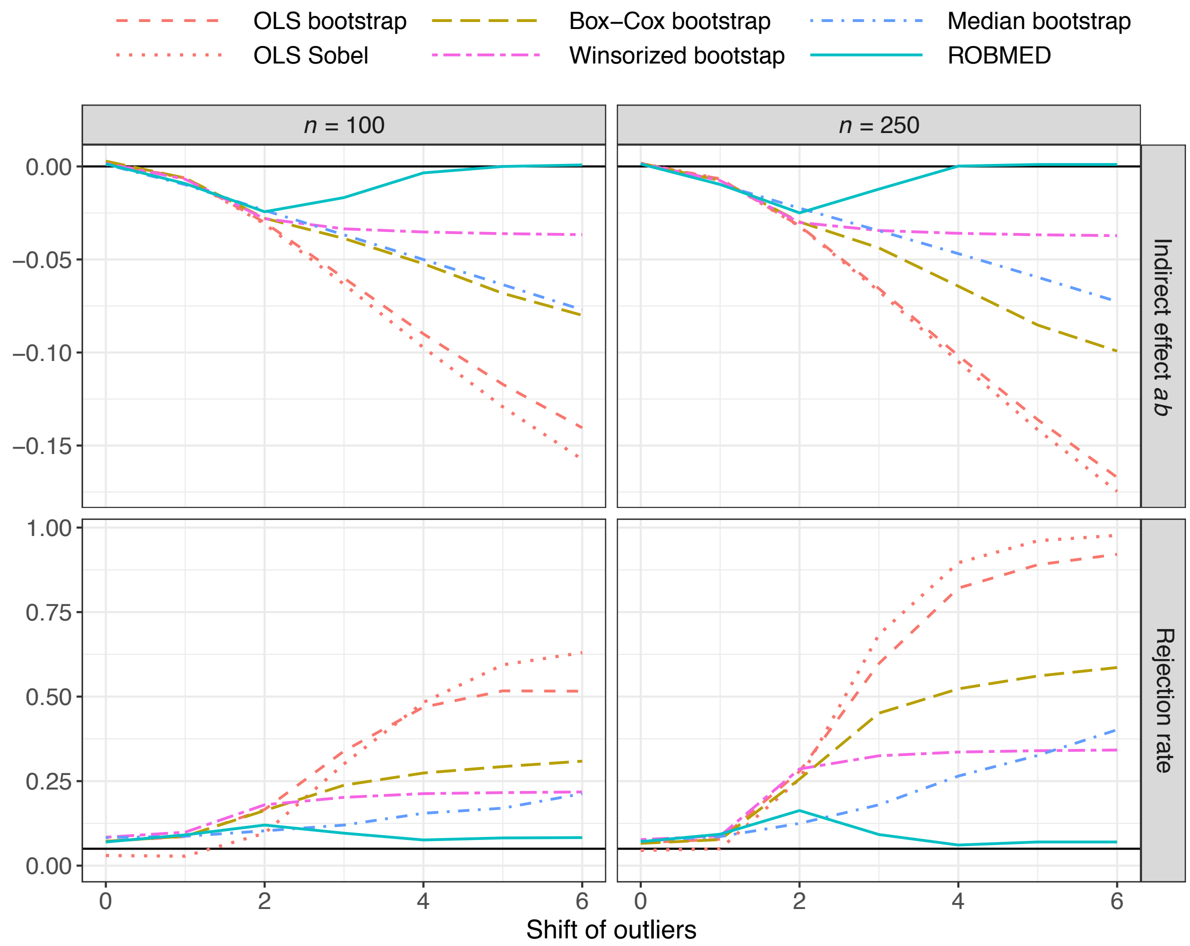

Figure 14. Results from 1000 simulation runs for simulation design 2 with varying distance of outliers (outlier probability 2%) and the setting with no mediation ($a = 0.4$, $b = 0$). The top row shows the average estimates of the indirect effect and includes a horizontal reference line for the true indirect effect $ab = 0$. The bottom row displays the rejection rate of the corresponding tests (i.e., the realized size), and a horizontal line is drawn for the nominal size $\alpha = 0.05$ (the closer to this line the better).

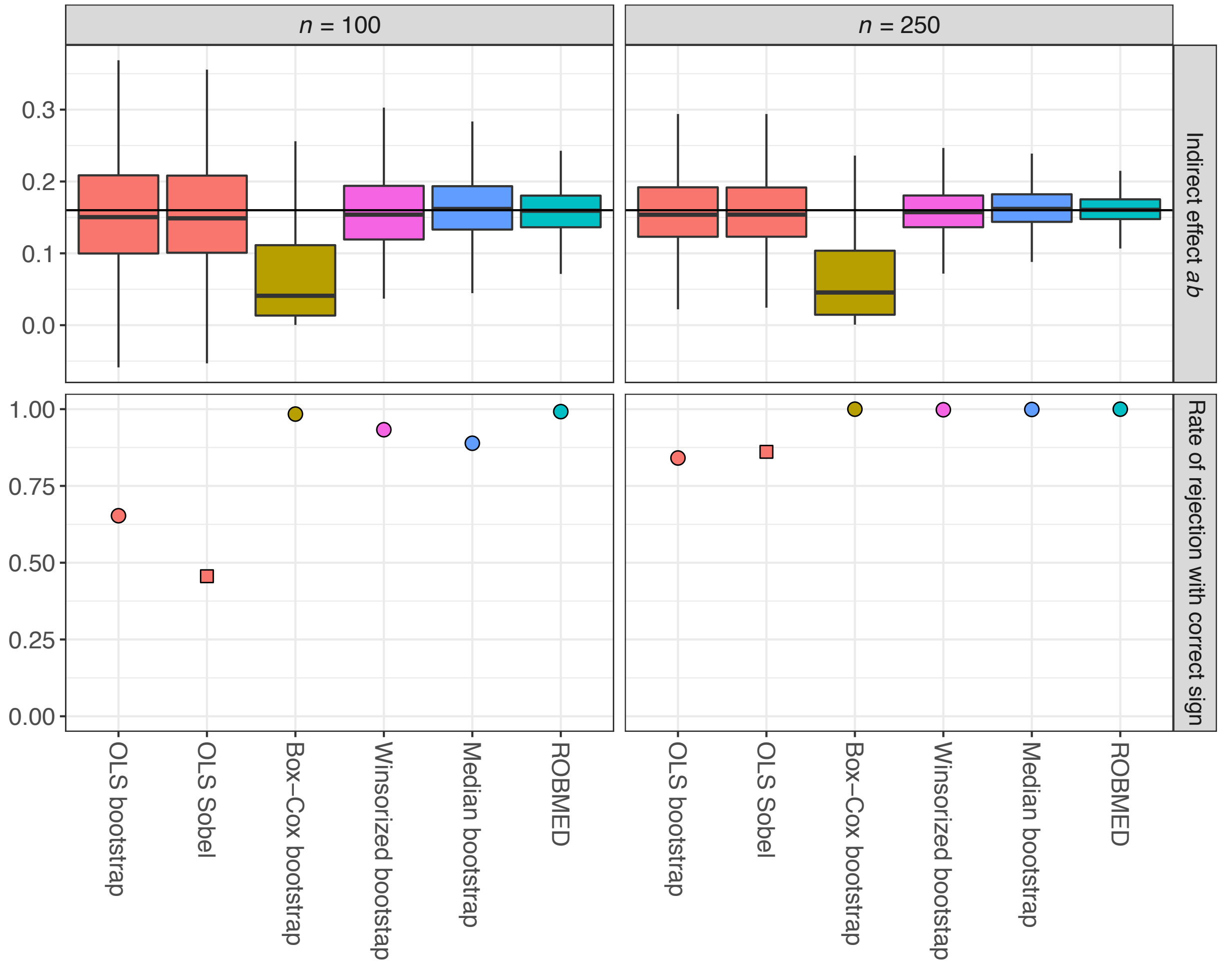

Figure 15. Results from 1000 simulation runs for simulation design 2 with centered log-normal distributions and the setting with mediation ($a = b = c = 0.4$). The top row shows the average estimates of the indirect effect and includes a horizontal reference line for the true indirect effect $ab = 0.16$. The bottom row displays the rate of how often the methods reject the null hypothesis and the corresponding estimate of *ab* has the correct sign (a measure of realized power of the tests in the presence of outliers; the higher this rate the better).

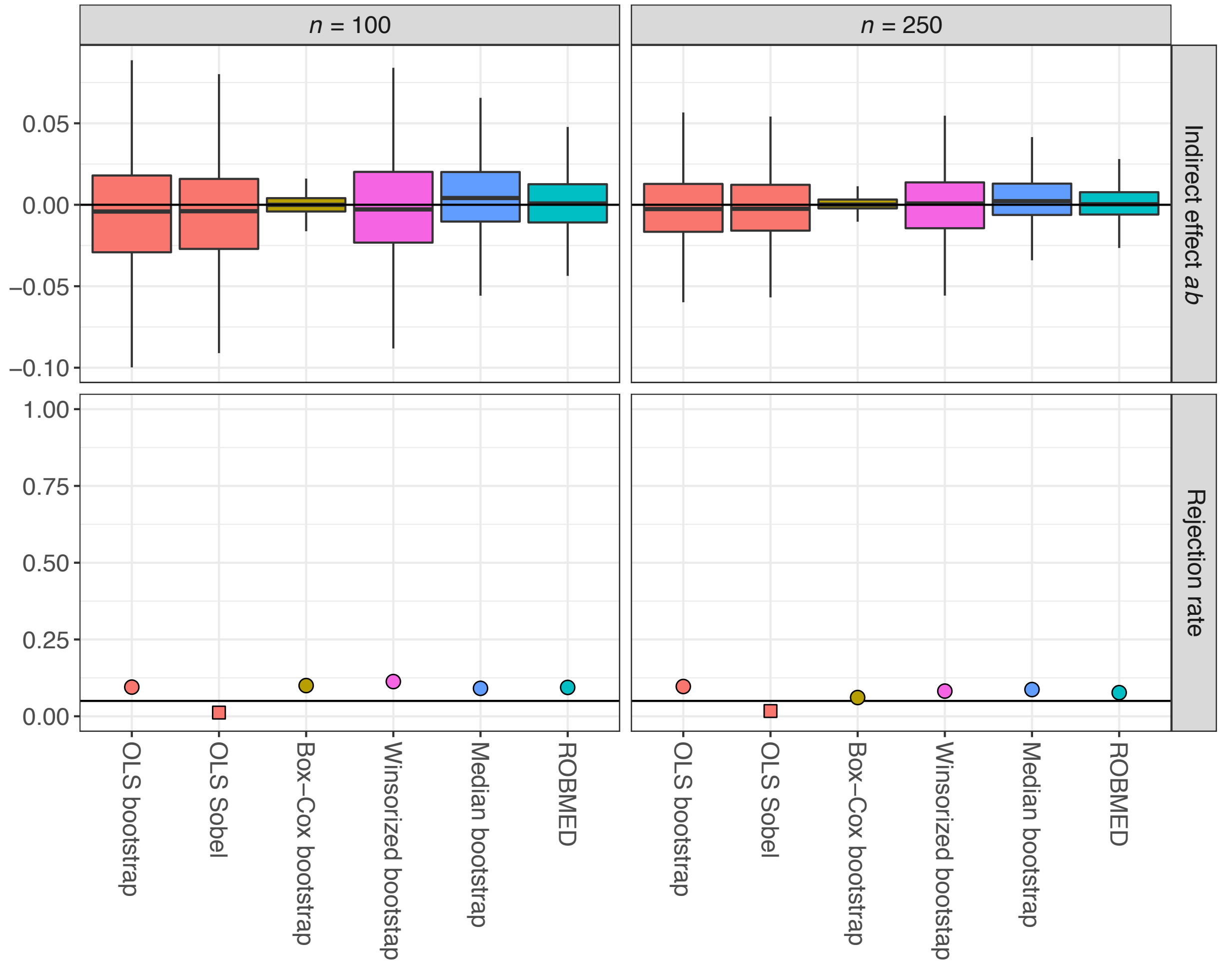

Figure 16. Results from 1000 simulation runs for simulation design 2 with centered log-normal distributions and the setting with mediation ($a = c = 0.4$, $b = 0$). The top row shows the average estimates of the indirect effect and includes a horizontal reference line for the true indirect effect $ab = 0$. The bottom row displays the rejection rate of the corresponding tests (i.e., the realized size), and a horizontal line is drawn for the nominal size $\alpha = 0.05$ (the closer to this line the better).

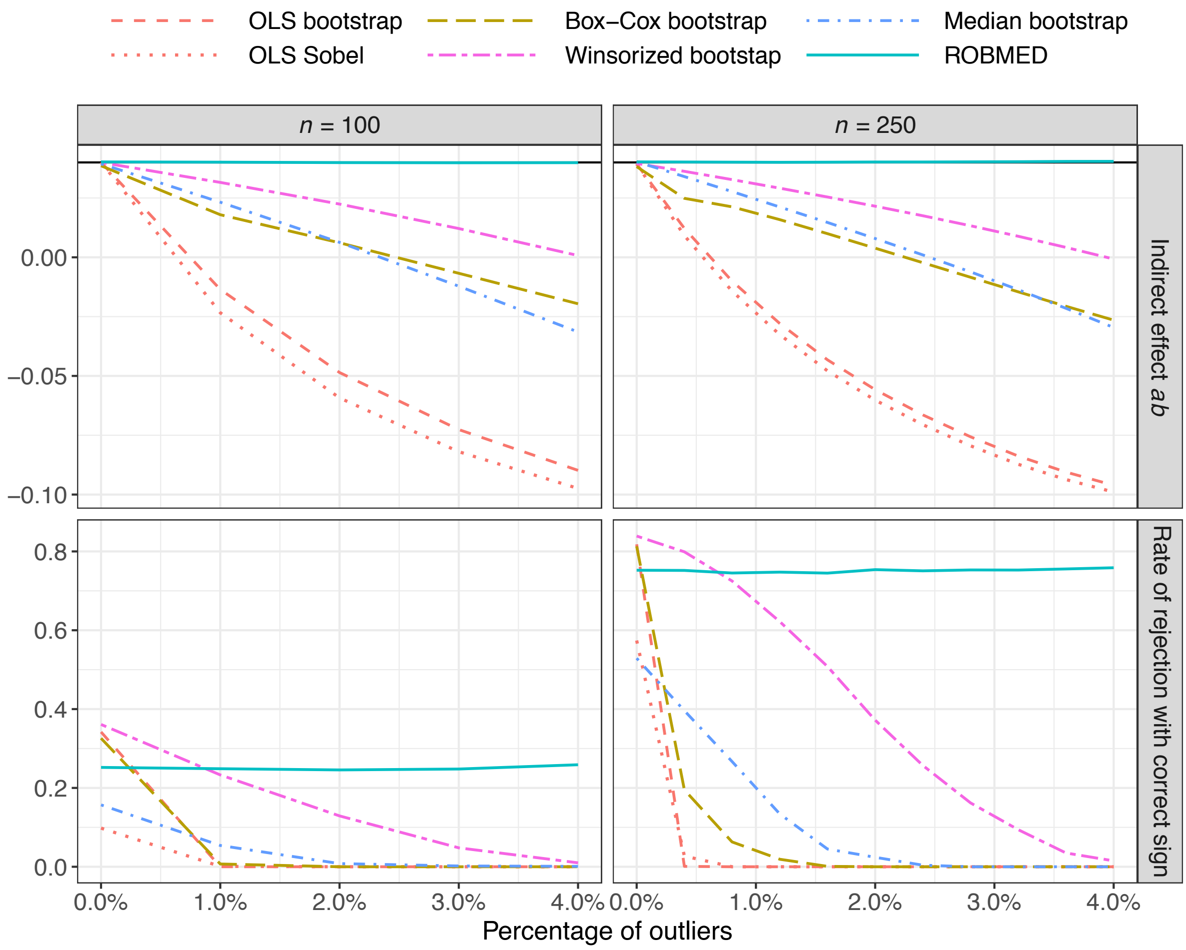

Figure 17. Results from 1000 simulation runs for simulation design 4 with varying percentage of outliers and the setting with mediation ($a = 0.2$, $b = 0.2$). The top row shows the average estimates of the indirect effect and includes a horizontal reference line for the true indirect effect $ab = 0.04$. The bottom row displays the rate of how often the methods reject the null hypothesis and the corresponding estimate of *ab* has the correct sign (a measure of realized power of the tests in the presence of outliers; the higher this rate the better).

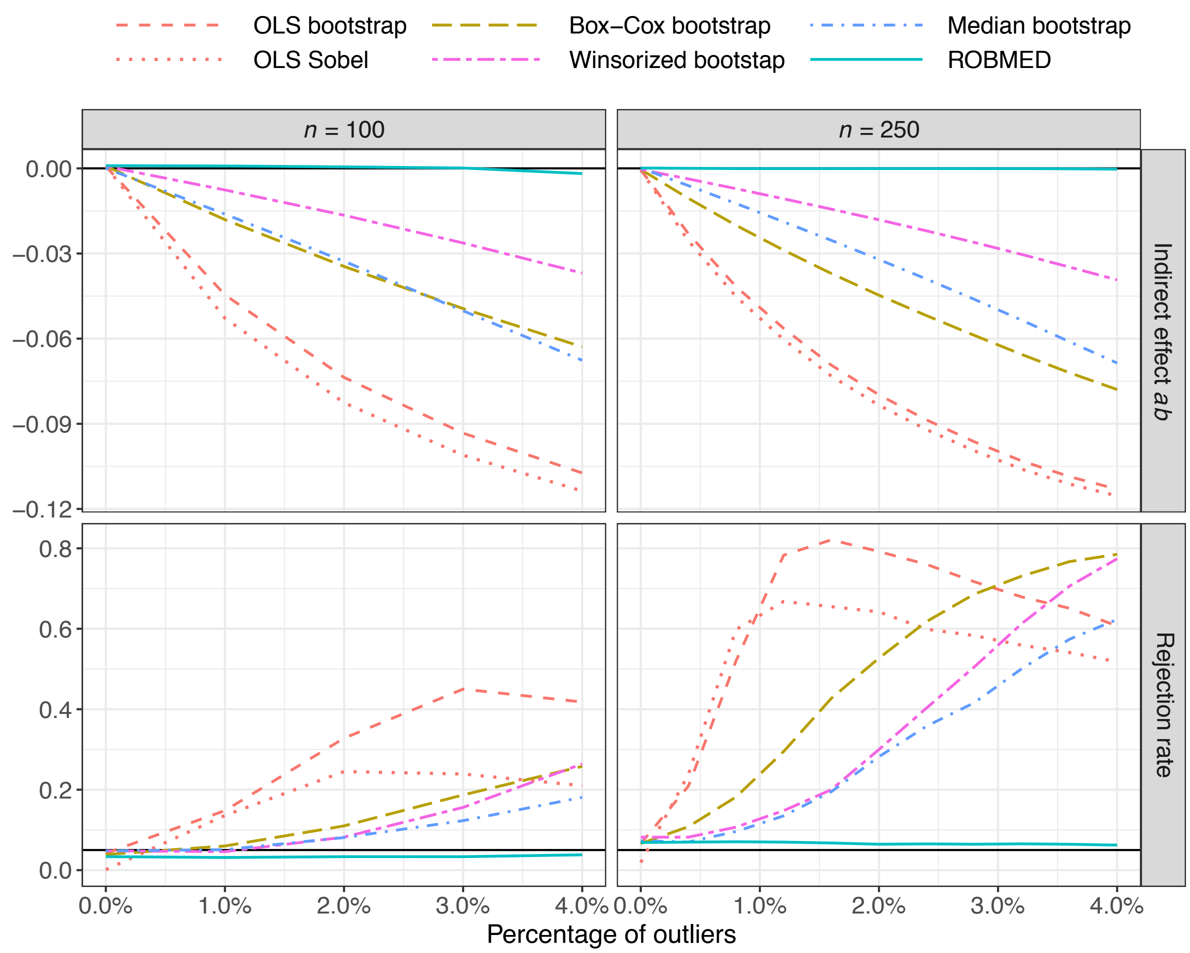

Figure 18. Results from 1000 simulation runs for simulation design 4 with varying percentage of outliers and the setting with no mediation ($a = 0.2$, $b = 0$). The top row shows the average estimates of the indirect effect and includes a horizontal reference line for the true indirect effect $ab = 0$. The bottom row displays the rejection rate of the corresponding tests (i.e., the realized size), and a horizontal line is drawn for the nominal size $\alpha = 0.05$ (the closer to this line the better).

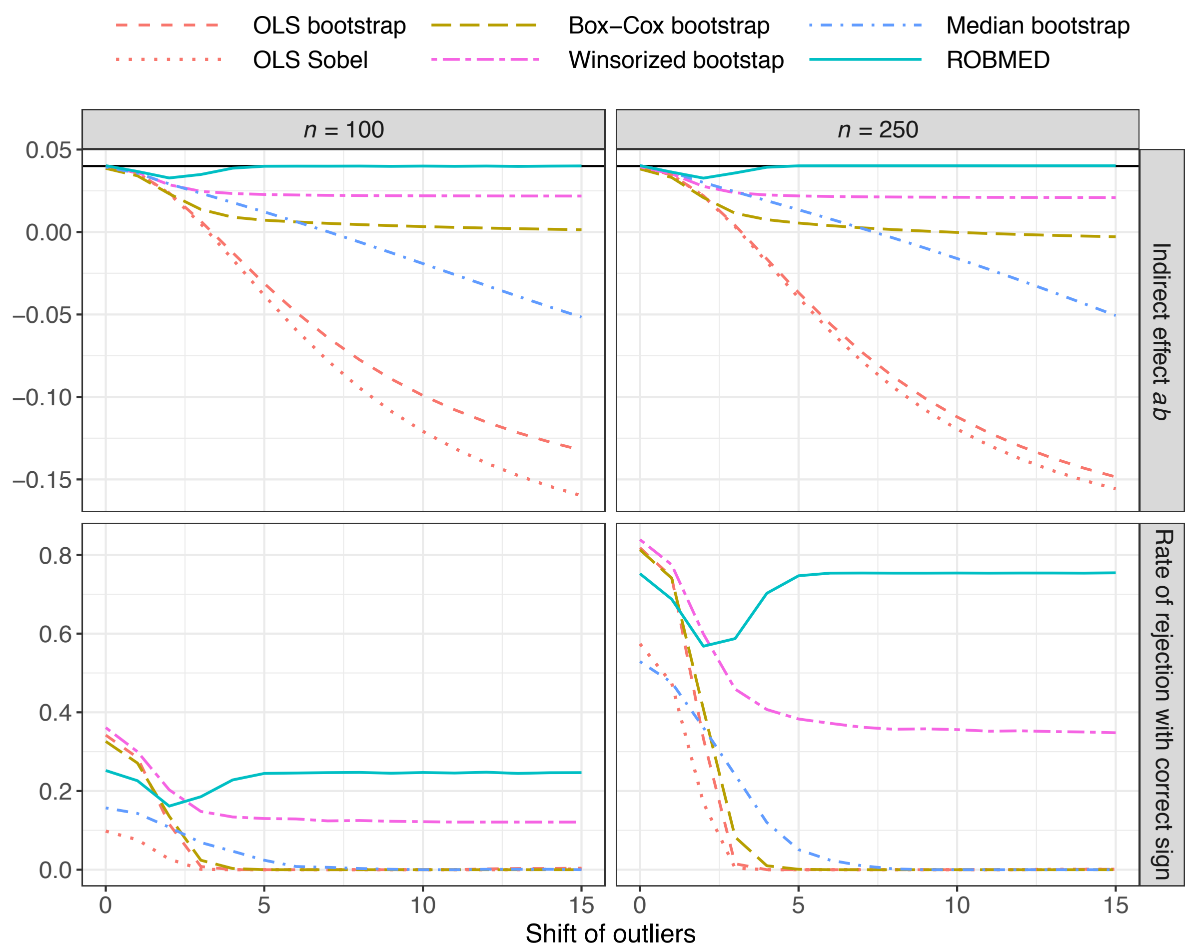

Figure 19. Results from 1000 simulation runs for simulation design 4 with varying distance of outliers and the setting with mediation ($a = 0.2$, $b = 0.2$). The top row shows the average estimates of the indirect effect and includes a horizontal reference line for the true indirect effect $ab = 0.04$. The bottom row displays the rate of how often the methods reject the null hypothesis and the corresponding estimate of ab has the correct sign (a measure of realized power of the tests in the presence of outliers; the higher this rate the better).

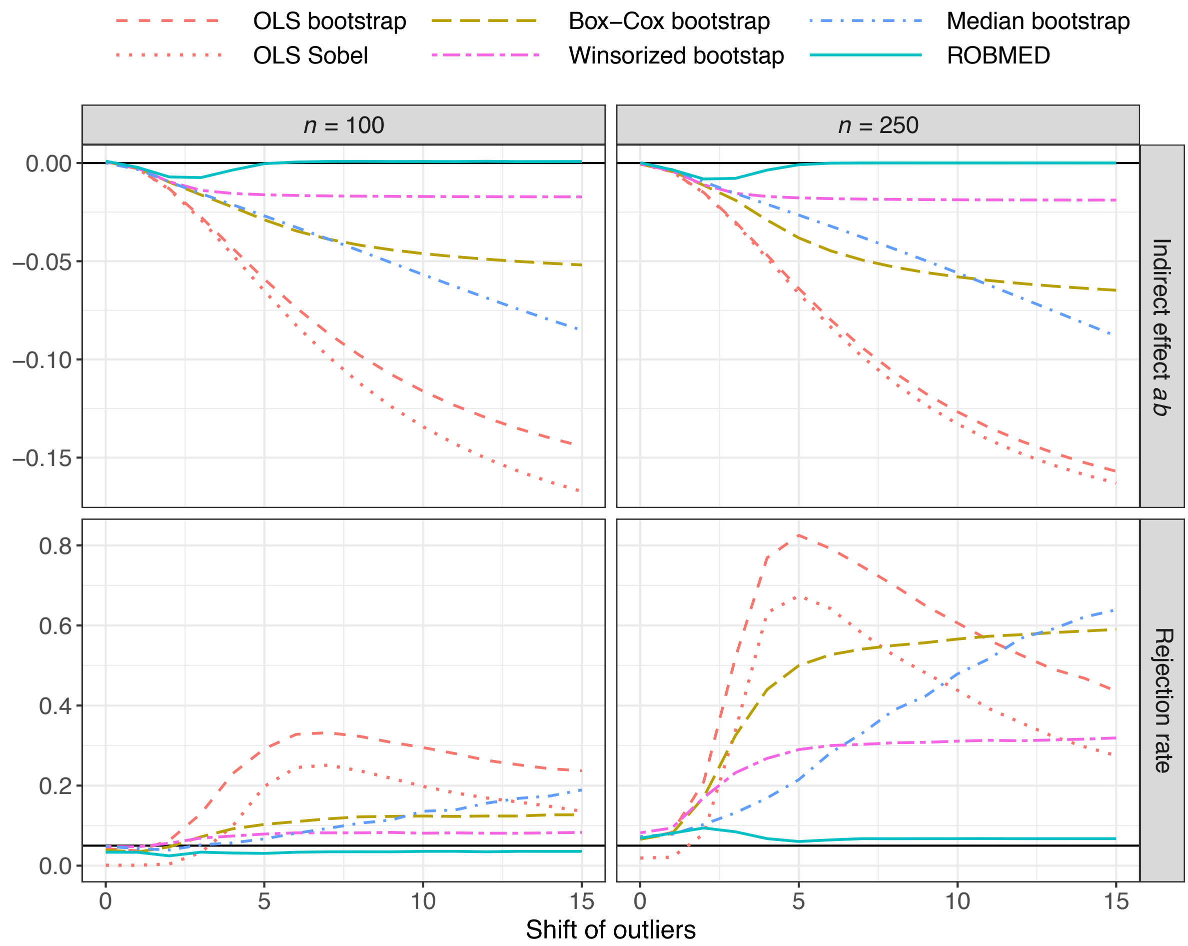

Figure 20. Results from 1000 simulation runs for simulation design 4 with varying distance of outliers and the setting with no mediation ($a = 0.2$, $b = 0$). The top row shows the average estimates of the indirect effect and includes a horizontal reference line for the true indirect effect $ab = 0$. The bottom row displays the rejection rate of the corresponding tests (i.e., the realized size), and a horizontal line is drawn for the nominal size $\alpha = 0.05$ (the closer to this line the better).

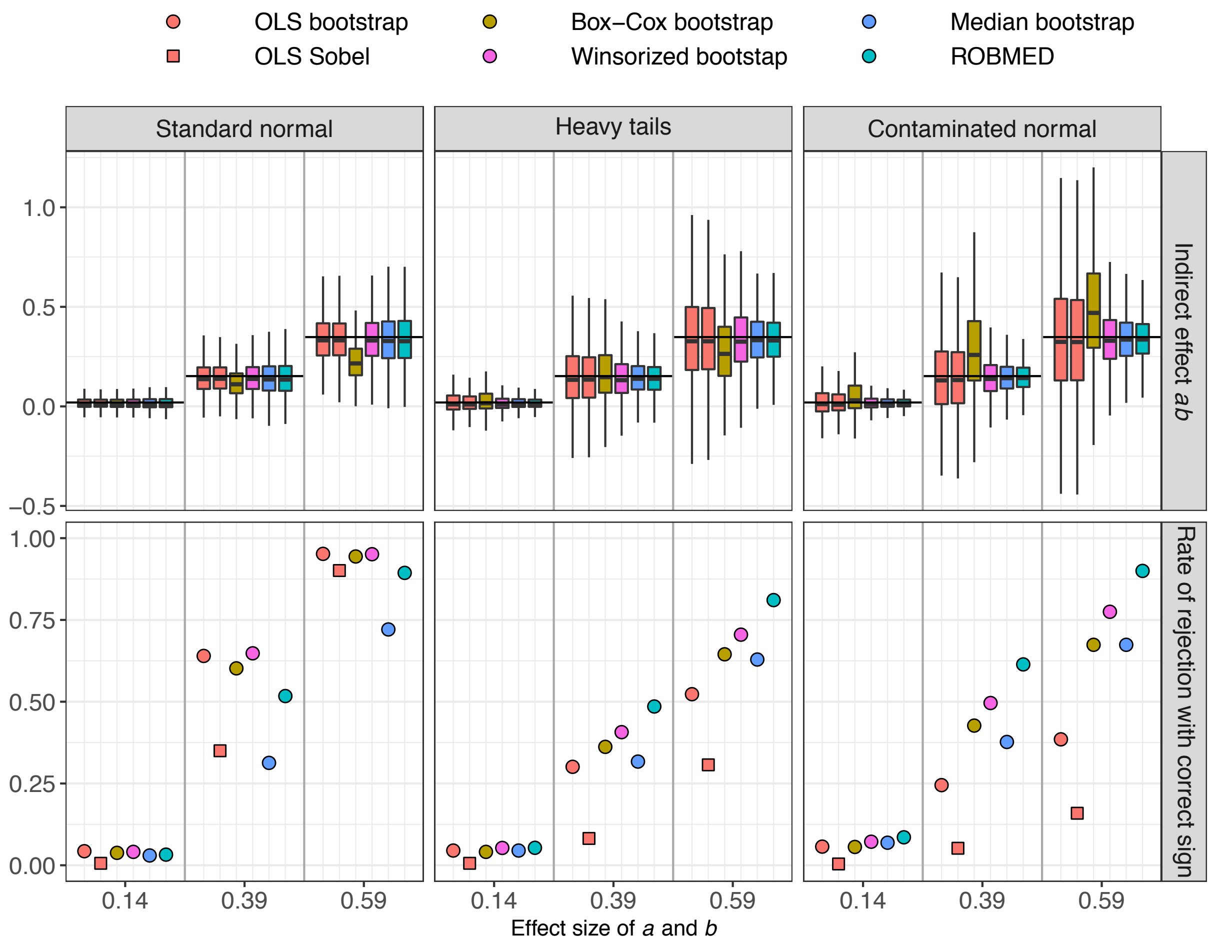

Figure 21. Results from 1000 simulation runs for simulation design 5 with different error distributions, the setting with mediation ($a = b = c = 0.14, 0.39, 0.59$), and sample size $n = 50$. The top row contains box plots of the estimates of the indirect effect, including horizontal reference lines for the true indirect effect ab. Points outside the whiskers are not displayed for better readability. The bottom row displays the rate of how often the methods reject the null hypothesis and the corresponding estimate of *ab* has the correct sign (a measure of realized power of the tests; the higher this rate the better). The columns correspond to the three considered distributions of the error terms.

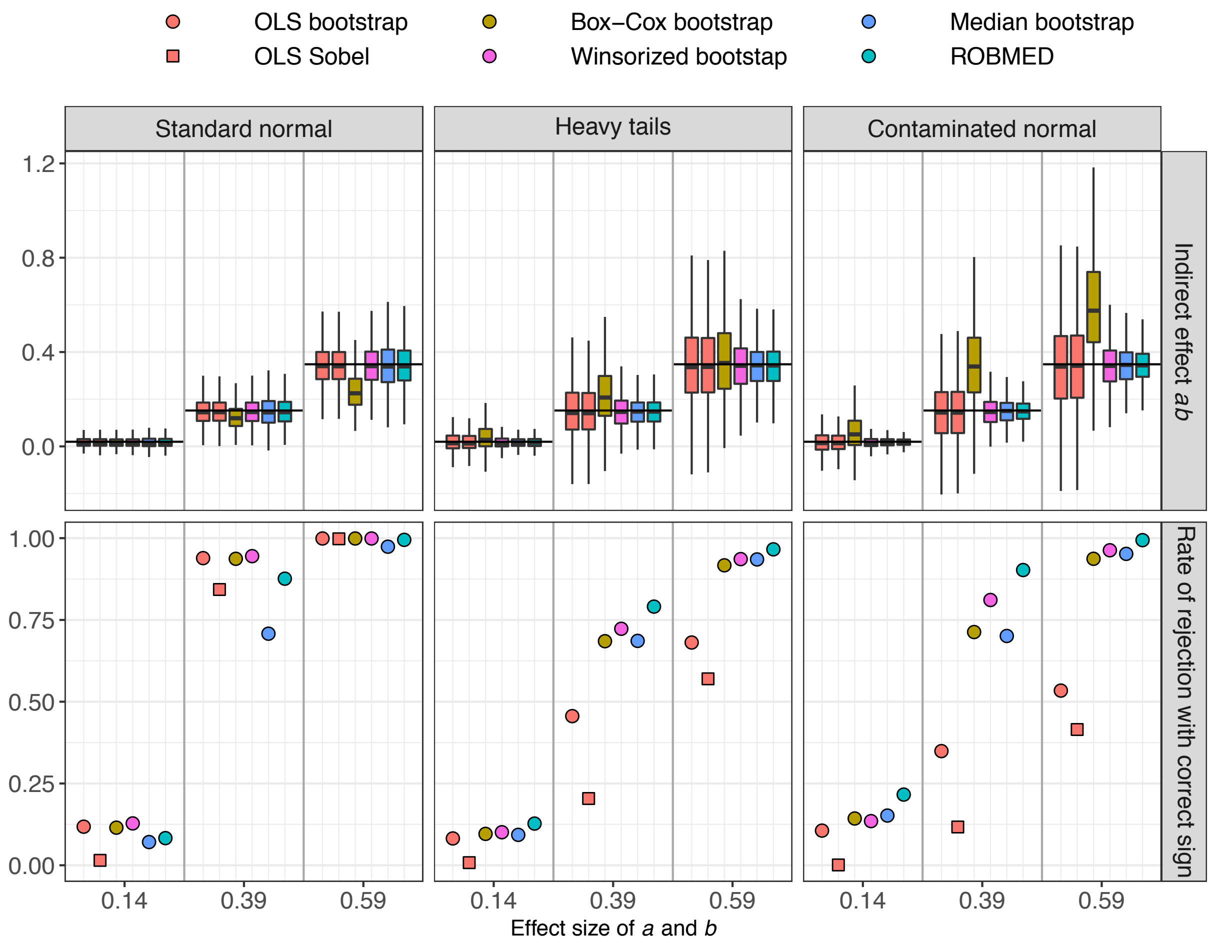

Figure 22. Results from 1000 simulation runs for simulation design 5 with different error distributions, the setting with mediation ($a = b = c = 0.14, 0.39, 0.59$), and sample size $n = 100$. The top row contains box plots of the estimates of the indirect effect, including horizontal reference lines for the true indirect effect ab. Points outside the whiskers are not displayed for better readability. The bottom row displays the rate of how often the methods reject the null hypothesis and the corresponding estimate of *ab* has the correct sign (a measure of realized power of the tests; the higher this rate the better). The columns correspond to the three considered distributions of the error terms.

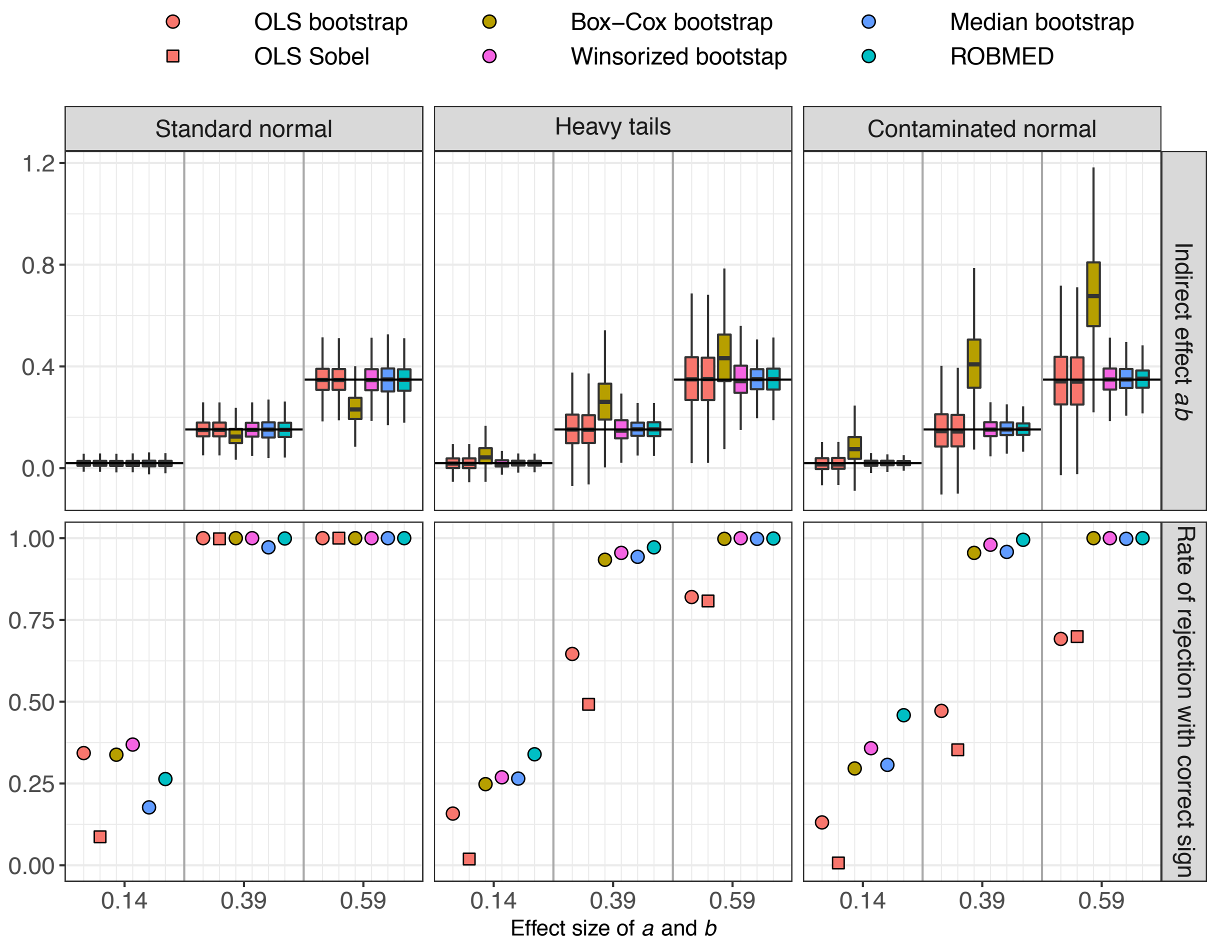

Figure 23. Results from 1000 simulation runs for simulation design 5 with different error distributions, the setting with mediation ($a = b = c = 0.14, 0.39, 0.59$), and sample size $n = 200$. The top row contains box plots of the estimates of the indirect effect, including horizontal reference lines for the true indirect effect ab. Points outside the whiskers are not displayed for better readability. The bottom row displays the rate of how often the methods reject the null hypothesis and the corresponding estimate of *ab* has the correct sign (a measure of realized power of the tests; the higher this rate the better). The columns correspond to the three considered distributions of the error terms.

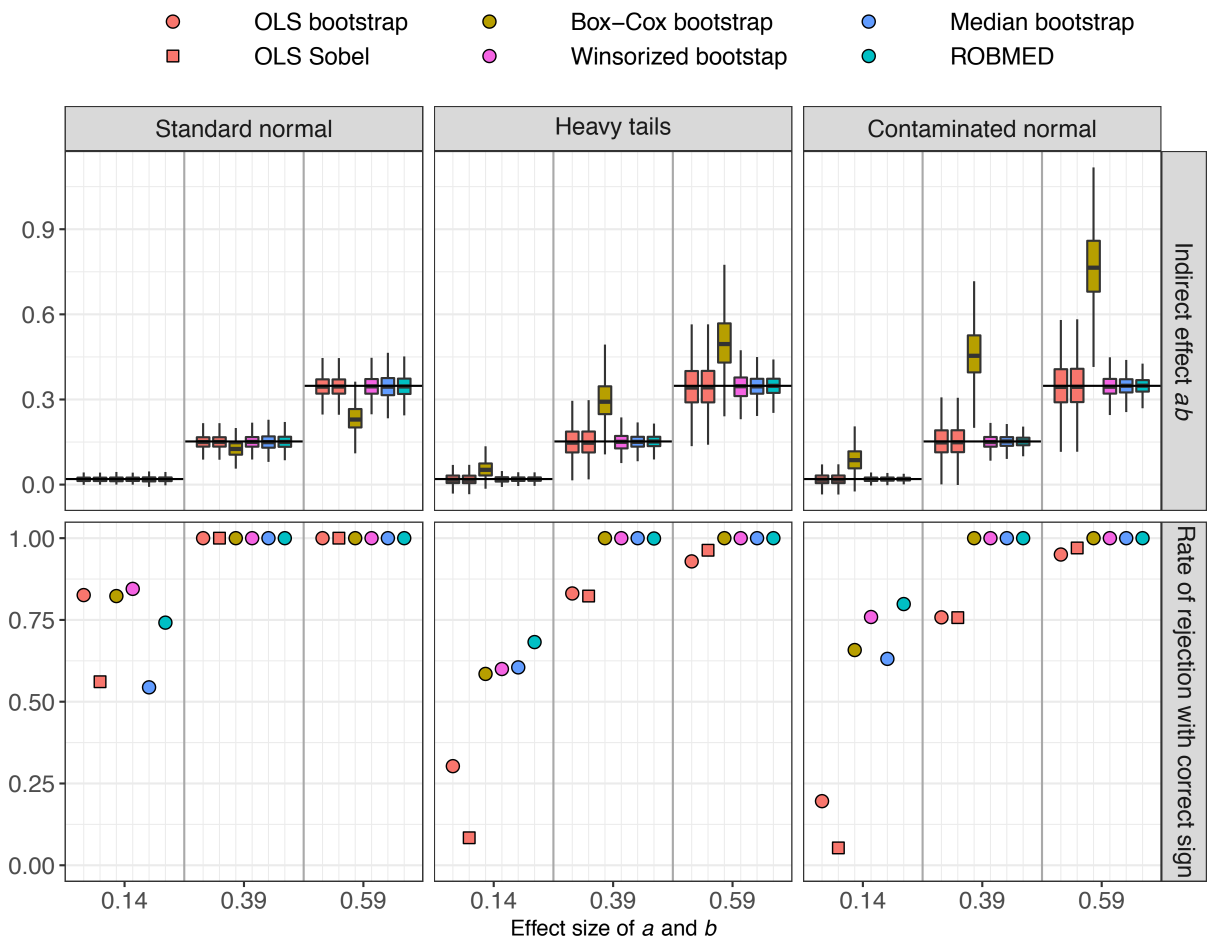

Figure 24. Results from 1000 simulation runs for simulation design 5 with different error distributions, the setting with mediation ($a = b = c = 0.14, 0.39, 0.59$), and sample size $n = 500$. The top row contains box plots of the estimates of the indirect effect, including horizontal reference lines for the true indirect effect ab. Points outside the whiskers are not displayed for better readability. The bottom row displays the rate of how often the methods reject the null hypothesis and the corresponding estimate of *ab* has the correct sign (a measure of realized power of the tests; the higher this rate the better). The columns correspond to the three considered distributions of the error terms.

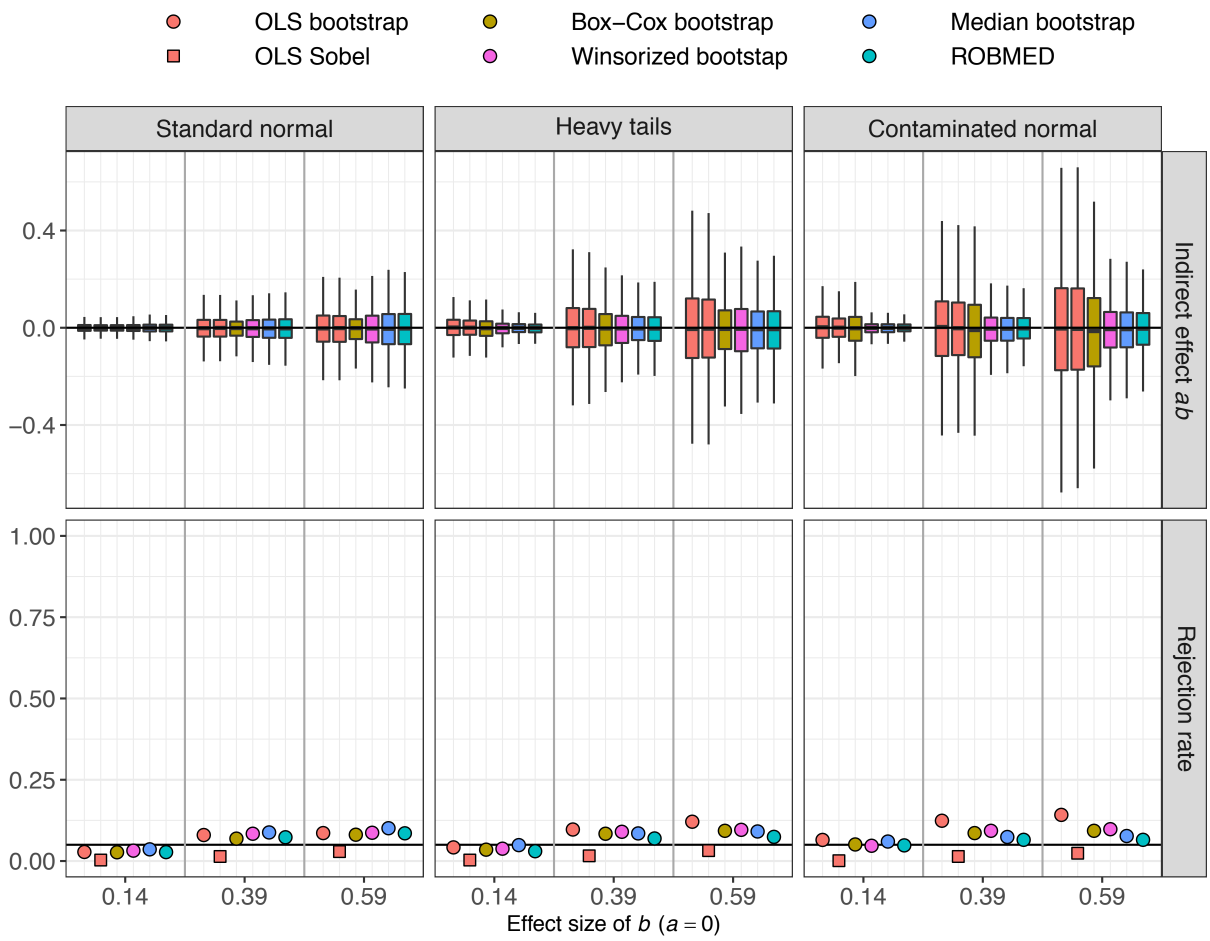

Figure 25. Results from 1000 simulation runs for simulation design 5 with different error distributions, the setting with no mediation ($a = 0$, $b = c = 0.14, 0.39, 0.59$), and sample size $n = 50$. The top row contains box plots of the estimates of the indirect effect, including a horizontal reference line for the true indirect effect $ab = 0$. Points outside the whiskers are not displayed for better readability. The bottom row displays the rejection rate of the corresponding tests (i.e., the realized size), and a horizontal line is drawn for the nominal size $\alpha = 0.05$ (the closer to this line the better). The columns correspond to the three considered distributions of the error terms.

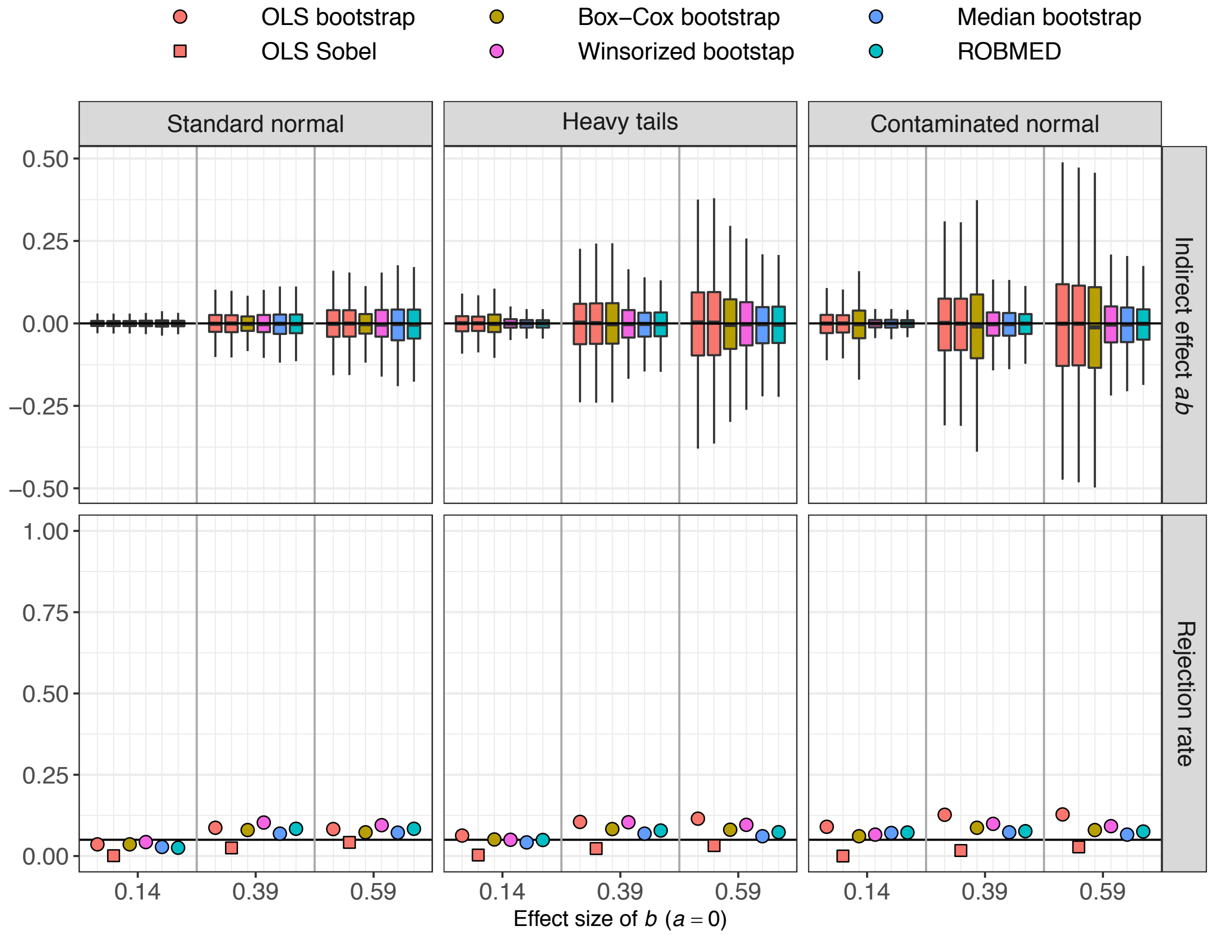

Figure 26. Results from 1000 simulation runs for simulation design 5 with different error distributions, the setting with no mediation ($a = 0$, $b = c = 0.14, 0.39, 0.59$), and sample size $n = 100$. The top row contains box plots of the estimates of the indirect effect, including a horizontal reference line for the true indirect effect $ab = 0$. Points outside the whiskers are not displayed for better readability. The bottom row displays the rejection rate of the corresponding tests (i.e., the realized size), and a horizontal line is drawn for the nominal size $\alpha = 0.05$ (the closer to this line the better). The columns correspond to the three considered distributions of the error terms.

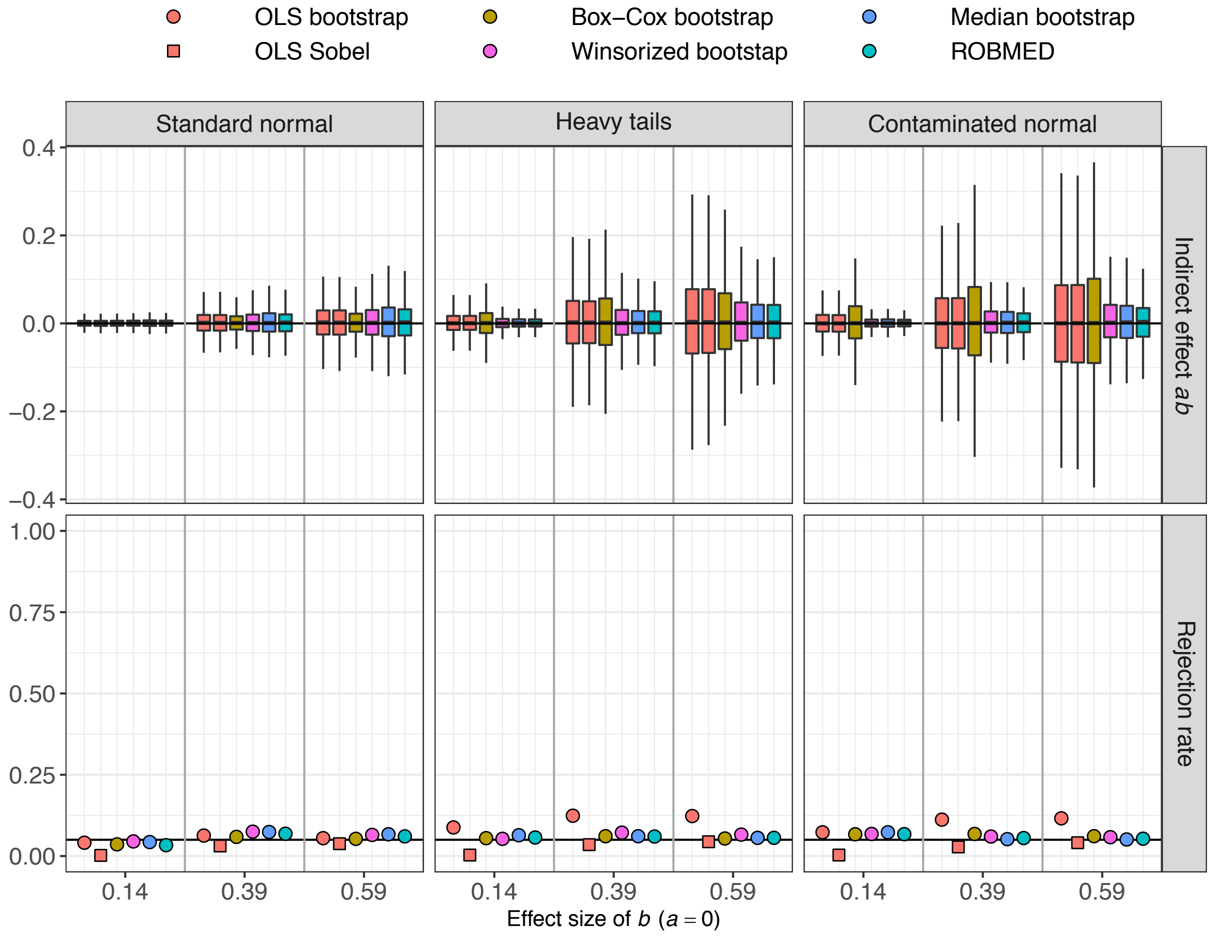

Figure 27. Results from 1000 simulation runs for simulation design 5 with different error distributions, the setting with no mediation ($a = 0$, $b = c = 0.14, 0.39, 0.59$), and sample size $n = 200$. The top row contains box plots of the estimates of the indirect effect, including a horizontal reference line for the true indirect effect $ab = 0$. Points outside the whiskers are not displayed for better readability. The bottom row displays the rejection rate of the corresponding tests (i.e., the realized size), and a horizontal line is drawn for the nominal size $\alpha = 0.05$ (the closer to this line the better). The columns correspond to the three considered distributions of the error terms.

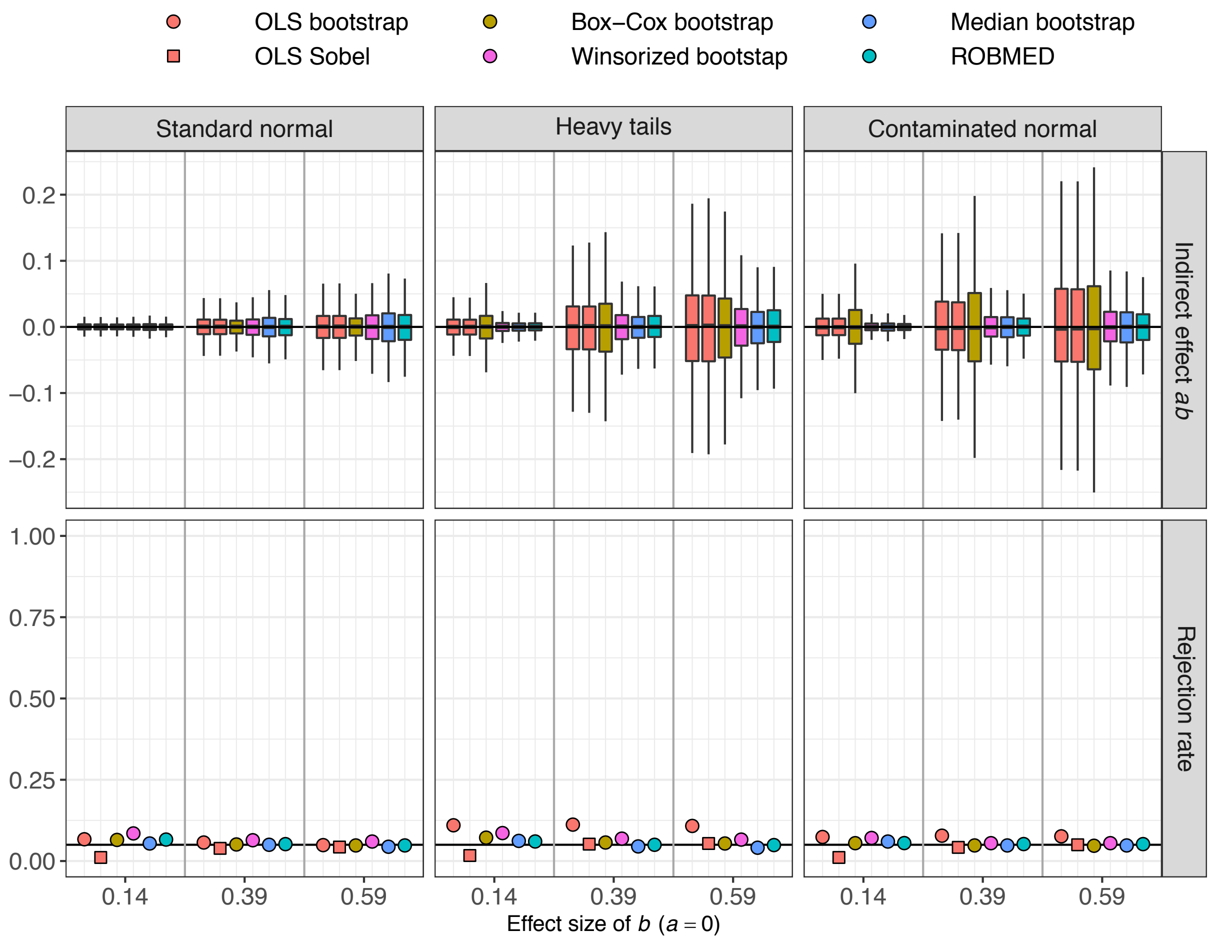

Figure 28. Results from 1000 simulation runs for simulation design 5 with different error distributions, the setting with no mediation ($a = 0$, $b = c = 0.14, 0.39, 0.59$), and sample size $n = 500$. The top row contains box plots of the estimates of the indirect effect, including a horizontal reference line for the true indirect effect $ab = 0$. Points outside the whiskers are not displayed for better readability. The bottom row displays the rejection rate of the corresponding tests (i.e., the realized size), and a horizontal line is drawn for the nominal size $\alpha = 0.05$ (the closer to this line the better). The columns correspond to the three considered distributions of the error terms.