Published in final edited form as:
Psychol Methods. 2014 March ; 19(1): 1–20. doi:10.1037/a0033820.

Robust Mediation Analysis Based on Median Regression

Ying Yuan and
Department of Biostatistics, The University of Texas MD Anderson Cancer Center

David P. MacKinnon
Department of Psychology, Arizona State University

Abstract

Mediation analysis has many applications in psychology and the social sciences. The most prevalent methods typically assume that the error distribution is normal and homoscedastic. However, this assumption may rarely be met in practice, which can affect the validity of the mediation analysis. To address this problem, we propose robust mediation analysis based on median regression. Our approach is robust to various departures from the assumption of homoscedasticity and normality, including heavy-tailed, skewed, contaminated, and heteroscedastic distributions. Simulation studies show that under these circumstances, the proposed method is more efficient and powerful than standard mediation analysis. We further extend the proposed robust method to multilevel mediation analysis, and demonstrate through simulation studies that the new approach outperforms the standard multilevel mediation analysis. We illustrate the proposed method using data from a program designed to increase reemployment and enhance mental health of job seekers.

Keywords

Mediation effect; Ordinary least squares; Maximum likelihood; Heteroscedastic; Heavy-tailed; Skewed; Outlier; Multilevel model

The aim of mediation analysis is to determine whether the relation between the independent variable and the dependent variable is due, wholly or in part, to a mediating variable that transmits the relation from the independent variable to the dependent variable (Baron & Kenny, 1986; MacKinnon, 2008). Mediation analysis has many applications in psychology, sociology, medicine and other fields. Since the seminal Baron and Kenny (1986) article, extensive research has been conducted on the accuracy of statistical mediation analysis. For example, MacKinnon and Dwyer (1993) investigated mediation analysis for binary outcomes. Collins, Graham and Flaherty (1998) provided an alternative framework for evaluating the mediation effect based on transitions to the values of the mediator and the outcome variable. Kraemer, Wilson, Fairburn and Agras (2002) outlined an analytic framework to identify mediators and moderators in randomized clinical trials. MacKinnon, Lockwood, Hoffman, West and Sheets (2002) compared fourteen methods to test the statistical significance of the mediation effect. Shrout and Bolger (2002) outlined a series of steps and emphasized the use of bootstrap approaches to assess mediation for small to moderate sample sizes (see also Bollen & Stine, 1990; Lockwood & MacKinnon, 1998).

In the context of multilevel modeling, Kenny, Kashy and Bolger (1998) identified two types of multilevel mediation, namely lower level and upper level mediation. In upper level

Correspondence concerning this article should be addressed to Ying Yuan, Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030. yyuan@mdanderson.org.

mediation, the independent variable for which the effect is mediated is an upper level variable, a variable appearing in the upper level of the model hierarchy; and in lower level mediation, the independent variable is a lower level variable, a variable appearing in the lower level of the model hierarchy. Krull and MacKinnon (1999, 2001) investigated various upper level mediation models and compared the performance of these models with that of single-level mediation models. Kenny, Korchmaros, and Bolger (2003) described a lower level mediation model in which each upper level unit has an individual mediation effect. Bauer, Preacher and Gil (2006) extended the work of Kenny et al. (2003), and proposed a method to yield consistent estimates of the variance components by simultaneously fitting the two mediation regression equations using a selection variable. Yuan and MacKinnon (2009) proposed a Bayesian approach for multilevel mediation analysis. For a comprehensive review of mediation analysis, see MacKinnon (2008).

The basic mediation framework, also known as the single-level mediation model, involves a three-variable system in which an independent variable causes a mediating variable, which in turn causes a dependent variable (Baron & Kenny, 1986; MacKinnon, 2008). Under this causality assumption, relations among the three variables can be expressed as three linear regression models; although only two of the three equations are required for the estimation of mediation. The standard mediation analysis fits these regression models using the ordinary least squares (OLS) or maximum likelihood method, and then, based on the estimates of regression parameters, draws statistical inference on the mediation effect. During this OLS-based estimation process, two standard distributional assumptions, homoscedasticity and normality, are often made on the errors of the regression models. Homoscedasticity means that the errors share a common variance, and the normality assumption means that the errors follow a normal distribution. When the homoscedasticity assumption is violated, the errors are called heteroscedastic.

Unfortunately, the assumptions of homoscedasticity and normality are rarely met in practice. Much empirical evidence suggests that heteroscedasticity exists in a wide variety of research areas (Grissom, 2000; Keppel & Wickens, 2004; Micceri, 1989; Tomarken & Serlin, 1986; Wilcox & Keselman, 2003). Keselman, Huberty, et al. (1998) reported that it is not uncommon in psychological studies to see a ratio larger than 8:1 when comparing the largest to the smallest variance across the different covariates, and in some extreme cases, to see a ratio as large as 566:1. Even in experiments that use randomization, regarded as the gold standard for comparing different treatments, the experimental variable may cause differences in variability between groups. Bryk and Raudenbush (1988) provided some examples of this, and Grissom and Kim (2005) explained why heteroscedasticity is so common in psychosocial data.

In addition to the occurrence of heteroscedasticity, the normality assumption is also frequently violated in psychosocial data. The violations of normality commonly encountered in practice include heavy tails, skewness, outliers, contamination and multimodality. Micceri (1989) examined 440 large data sets from the psychological and educational literature, including 125 psychometric measures, such as scales measuring personality, anxiety, and satisfaction. None of the data sets were found to be normally distributed; instead, the distributions were frequently heavy-tailed and skewed. Based on Hogg's (1974) Q measure and the C ratio of Elashoff and Elashoff (1978), Micceri (1989) classified these distributions into five categories according to their tail weights or asymmetry: uniform, below Gaussian, moderate contamination, extreme contamination, and double exponential. Of the 125 psychometric measures, about 23.3% belonged to the double exponential category, 28% had extreme contamination, and only approximately 13.6% were "about Gaussian." Micceri's findings were based on the marginal distributions of the psychometric variables, which do not necessarily mean that the errors violated the normality assumption when these variables were used as dependent variables in regression. Nevertheless, in practice the violation of normality in the marginal distribution often provides evidence of the non-normality of the errors because regressing (or conditioning) on some independent variables typically cannot completely remove the non-normality of the dependent variables.

In the presence of non-normality and heteroscedasticity, the OLS method, and thus the standard OLS-based mediation analysis, may break down and contribute to misleading results, including inefficient estimates, invalid confidence intervals (CIs), distorted type I error rates, and low power (Erceg-Hurn & Mirosevich, 2008; Grissom & Kim, 2001; Keselman, Algina, Lix, Wilcox & Deering, 2008; Serlin & Hartwell, 2004; Wilcox, 1998; among others). Data transformations, such as the square root or logarithm transformation, are commonly used to bring the errors in estimation closer in line with the assumption of normality and homoscedasticity. However, data transformation often fails to achieve normality and homoscedasticity and may not adequately deal with outliers. In addition, data transformation causes difficulties in interpreting analytic results, as the results are based on the transformed scale rather than the original scale (Grissom, 2000; Lix, Keselman, & Keselman, 1996).

In the context of mediation analysis, Bollen and Stine (1990) provided empirical evidence that the estimate of a mediation effect could be substantially sensitive to outliers. Finch, West and MacKinnon (1997) showed that under non-normality, the empirical standard error (SE) of the standard maximum likelihood estimate of the mediation effect was overestimated, and the information-based SE estimate was negatively biased. Based on simulation studies, Lockwood (1999) also showed that the information-based SE estimate was negatively biased. Recently, Zu and Yuan (2010) proposed using local influence methods to identify observations that strongly affect the testing of mediation. They also proposed using robust methods, such as M-estimators, for parameter estimation and hypothesis testing of the mediation effect. In this article, we propose a robust mediation analysis method based on median regression. The proposed method is robust to various departures from normality and homoscedasticity, including heavy tails, skewness, outliers, and distributional contaminations. Our simulation studies show that when errors are heteroscedastic and/or non-normal, the proposed approach is more efficient and powerful than the standard mediation analysis for both single-level and multilevel mediations.

Single-level Mediation Model

Let y_i denote the measurement of the dependent (or outcome) variable, x_i denote the measurement of the independent variable, and m_i denote the measurement of the mediating variable (or mediator) for the ith subject, where i=1, …, n. Under the causality assumption that the independent variable causes the mediating variable, which in turn causes the dependent variable (see Figure 1), the single-level mediation model can be expressed in the form of three regression equations (Baron & Kenny, 1986):

$$y_i=\beta_{01}+\tau x_i+\varepsilon_{1i} \quad (1)$$

$$m_i=\beta_{02}+\alpha x_i+\varepsilon_{2i} \quad (2)$$

$$y_i=\beta_{03}+\beta m_i+\tau' x_i+\varepsilon_{3i} \quad (3)$$

where τ quantifies the relation between the independent variable and dependent variable; α measures the relation between the independent variable and mediating variable; β quantifies the relation between the mediating variable and dependent variable after adjusting for the effects of the independent variable; and τ' quantifies the relation between the independent variable and dependent variable after adjusting for the effect of the mediating variable. The regression equations 1 – 3 are not independent because equation 1 can be derived by substituting equation 2 into equation 3. Therefore, for the purpose of mediation analysis, we need to focus on only two equations, 2 and 3.

Under the single-level mediation model, the mediation effect can be defined as two equivalent forms: $\alpha\beta$ or $\tau - \tau'$ (MacKinnon, Warsi, & Dwyer, 1995). To estimate the mediation effect, we first obtain estimates of the unknown regression parameters that appeared in the mediation equations using the OLS method, and then estimate the mediation effect by $\hat{\alpha}\hat{\beta}$ or $\hat{\tau} - \hat{\tau}'$, where $\hat{\alpha}$, $\hat{\beta}$, $\hat{\tau}$ and $\hat{\tau}'$ are the OLS estimates. In this article, we focus on estimating $\alpha\beta$ as the mediation effect, but the proposed new methodology applies to estimating $\tau - \tau'$ as well. Note that the mediation analysis based on the mediation regression equations we consider here, i.e., the framework of Baron and Kenny (1986), does not provide confirming evidence for the causality of the mediation effect, but provides only the descriptive or explanatory information about the relationship among the independent variable, dependent variable and mediating variable. Establishing the causality of the mediation effect requires some additional assumptions of no unmeasured confounders (VanderWeele & Vansteelandt, 2009) and a different statistical inferential framework, e.g., the counterfactual causal model (Rubin, 1974).

In addition to the point estimate $\hat{\alpha}\hat{\beta}$, the confidence interval (CI) and test of the mediation effect are often of interest. The sampling distribution of $\hat{\alpha}\hat{\beta}$ is not normal (Lomnicki, 1967; MacKinnon et al., 2002). Several methods have been proposed to account for this fact when constructing the CI of $\hat{\alpha}\hat{\beta}$. One such method is the distribution of the product method (MacKinnon, Lockwood & Williams, 2004), which approximates the sampling distribution of $\hat{\alpha}\hat{\beta}$ with the product of two normal random variables with means equal to $\hat{\alpha}$ and $\hat{\beta}$ and variances equal to $\hat{\sigma}^2_{\hat{\alpha}}$ and $\hat{\sigma}^2_{\hat{\beta}}$, respectively, where $\hat{\sigma}^2_{\hat{\alpha}}$ and $\hat{\sigma}^2_{\hat{\beta}}$ denote the estimates of $\sigma^2_{\hat{\alpha}}$ and $\sigma^2_{\hat{\beta}}$ (i.e., the sampling variances of $\hat{\alpha}$ and $\hat{\beta}$). Meeker, Cornwell and Aroian (1981) provided critical values of the distribution of the product of two normal random variables that can be used to construct CIs. Alternatively, these critical values can be obtained on the basis of the empirical distribution of the product of two normal random variables through Monte Carlo simulation (Bauer et al., 2006; MacKinnon et al., 2004). Let c_{lower} and c_{upper} denote critical values corresponding to the lower and upper bounds of the CI, then the CI of the mediation effect is given by $(\hat{\alpha}\hat{\beta} + c_{\text{lower}} \times \hat{\sigma}_{\hat{\alpha}\hat{\beta}}, \hat{\alpha}\hat{\beta} + c_{\text{upper}} \times \hat{\sigma}_{\hat{\alpha}\hat{\beta}})$. The major advantage of the distribution of the product method is that it uses the standard outputs of linear regression (i.e., $\hat{\alpha}$, $\hat{\beta}$, $\hat{\sigma}^2_{\hat{\alpha}}$ and $\hat{\sigma}^2_{\hat{\beta}}$) and requires fitting mediation models only once. Therefore, the distribution of the product method is particularly appealing when the model fitting is time-consuming, e.g., in the case of complicated multilevel mediation models. Obviously, one limitation of the distribution of the product method is that it requires valid estimates of $\hat{\sigma}^2_{\hat{\alpha}}$ and $\hat{\sigma}^2_{\hat{\beta}}$ to make correct inference.

Another general approach to constructing the CI without imposing the normality assumption on $\hat{\alpha}\hat{\beta}$ is the bootstrap method based on resampling (Bollen & Stine, 1990; Efron, 1979). Compared to the distribution of the product method, the bootstrap method is more general and robust in the sense that it does not require the estimates of $\sigma^2_{\hat{\alpha}}$ and $\sigma^2_{\hat{\beta}}$. However, the bootstrap method is computationally intensive and requires repeatedly fitting the mediation models for each bootstrap sample, which could be a concern for complicated mediation models. In order to construct the CI, the bootstrap method repeatedly resamples the original data set with replacement, and then estimates the mediation effect for each of the bootstrap samples. The resulting collection of estimates forms the empirical distribution of the mediation effect, which can be used to construct the CI. In the simplest form of bootstrapping, called the percentile bootstrap, the $100 \times (1 - \alpha)\%$ CI of the mediation effect is given by $(q_{\alpha/2}, q_{1-\alpha/2})$, where $q_{\alpha/2}$ and $q_{1-\alpha/2}$ denote the $\alpha/2$th and $(1-\alpha/2)$th percentiles of the empirical distribution of the mediation effect. Other forms of bootstrapping that aim to improve its performance have been proposed. One variation that is important for mediation analysis is the bias-corrected bootstrap (Efron, 1987). The bias-corrected bootstrap adjusts and removes the potential estimation bias that arises because the true parameter value is not the median of the distribution of the bootstrap estimates, thereby in general yielding more accurate CIs than the percentile bootstrap when the mediation effect is nonzero (MacKinnon et al., 2004). Various R and SAS routines (MacKinnon et al., 2004), as well as SPSS macros (Hayes, 2012), are readily available to calculate bootstrap confidence intervals for mediation effects.

Problems with the OLS-based Mediation Analysis

Under the single-level mediation model, the standard mediation analysis is based on the OLS method, which estimates unknown parameters by minimizing the sum of the squared errors. For example, in equations 2 and 3, the OLS estimates of α and β are those that minimize

$$\sum_{i=1}^{n}(m_i - \hat{\beta}_{02} - \hat{\alpha}x_i)^2 \quad (4)$$

and

$$\sum_{i=1}^{n}(y_i - \hat{\beta}_{03} - \hat{\beta}m_i - \hat{\tau}'x_i)^2, \quad (5)$$

respectively. In principle, the OLS method does not require the normality and homoscedasticity assumptions to yield unbiased point estimates of the regression parameters. However, in order to construct the CI and conduct hypothesis testing in a valid and efficient way, homoscedasticity and/or normality assumptions of ε_{2i} and ε_{3i} are often required, depending on the specific statistical method used.

Specifically, the homoscedasticity assumption is required for the distribution of the product method, but is not essential for the bootstrap method. This is because the distribution of the product method requires the homoscedasticity assumption to ensure the validity of the OLS estimates of $\sigma^2_{\hat{\alpha}}$ and $\sigma^2_{\hat{\beta}}$; whereas the bootstrap method does not depend on the estimates of $\sigma^2_{\hat{\alpha}}$ and $\sigma^2_{\hat{\beta}}$. When the homoscedasticity assumption is violated, the OLS estimates of $\sigma^2_{\hat{\alpha}}$ and $\sigma^2_{\hat{\beta}}$ are biased and thus the distribution of the product method is invalid. To see the bias, consider a general linear regression model in the matrix form

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (6)$$

where $\boldsymbol{Y}$ is the vector of the dependent variable, $\boldsymbol{X}$ is the design matrix, including all independent variables, β is the vector of unknown regression coefficients, and ε is the vector of errors with a covariance matrix $\text{Var}(\varepsilon) = \sigma^2\boldsymbol{V}$. Under the OLS method, the estimate of β is $\hat{\beta} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$ and its associated sampling variance is

$$\text{Var}(\hat{\beta}) = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\sigma^2. \quad (7)$$

However, the true sampling variance of $\hat{\beta}$ is

$$\mathrm{Var}(\hat{\beta})=(\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\sigma^2, \quad (8)$$

which differs from the OLS estimate (7) when the errors are heteroscedastic (i.e., $\boldsymbol{V}$ is not an identity matrix). The bias of the OLS estimate of $\mathrm{Var}(\hat{\beta})$ can be addressed by using heteroscedasticity consistent (HC) covariance estimates, which remain consistent in the presence of heteroscedasticity of an unknown form. Long and Ervin (2000) compared several HC covariance estimates and recommended the following so-called HC3 estimate for general use:

$$\mathrm{HC3}=(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\mathrm{diag}\left[\frac{e_i^2}{(1-h_{ii})^2}\right]\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}, \quad (9)$$

where $e_i = y_i - \boldsymbol{x}_i\hat{\beta}$ and $h_{ii}=\boldsymbol{x}_i(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i^T$ with $\boldsymbol{xi}$ being the ith row of $\boldsymbol{X}$. Hayes and Cai (2007) provide SPSS and SAS macros to calculate HC covariance estimates.

Although the normality assumption is not essential for the validity of the bootstrap and the distribution of the product methods (at least for large samples), it is important for the efficiency of these methods. In particular, when the errors are normal (and homoscedastic), the OLS estimates are the maximum likelihood estimates and achieve a well-known optimality. Research shows that violations of the normality assumption, such as heavy tails, skewness, outliers and distributional contaminations, often result in inefficient estimates of α and β with large variances (Schrader & Hettmansperger, 1980), which often translate into low statistical power for testing mediation effects. The goal of the bootstrap and the distribution of the product methods is to reflect the true uncertainty of $\hat{\alpha}\hat{\beta}$. When the uncertainty of $\hat{\alpha}\hat{\beta}$ is in fact high (i.e., $\hat{\alpha}\hat{\beta}$ is an inefficient estimate), for example, caused by non-normality, neither the bootstrap nor the distribution of the product methods can address the low efficiency of the estimate, as we demonstrate in stimulation studies in a subsequent section.

Median Regression

Mean regression models, such as equations 1 to 3, describe how the mean or central location of the dependent variable changes with independent variables. However, the mean is not always an appropriate summary of the data distribution. It is well known that the mean is sensitive to outliers and performs poorly when distributions are skewed or heavy-tailed (Hill & Dixon, 1982; Wegman & Carroll, 1977). In these cases, the median provides a better summary of the central location of the data distribution.

Median regression models describe how the median of the dependent variable changes with the independent variables. As an example, consider a median regression of the mediating variable m_i on the independent variable x_i, which can be expressed in a form similar to equation 2, as follows:

$$m_i=\beta_{02}+\alpha x_i+e_{2i}. \quad (10)$$

However, unlike the mean regression model 2, which assumes $E(\varepsilon_{2i}|x_i) = 0$ and concerns the conditional mean function $E(m_i|x_i) = \beta_{02} + \alpha x_i$, the median regression assumes $M(e_{2i}|x_i) = 0$ and models the conditional median function of m_i as $M(m_i|x_i) = \beta_{02} + \alpha x_i$, where $M(\cdot \mid \cdot)$ denotes the conditional median. Nevertheless, the median regression models have a well-defined interpretation for covariate effects that is similar to that of the models based on the means in the sense that $M(m_i|x_i + 1) - M(m_i|x_i) = M(m_j|x_j + 1) - M(m_j|x_j)$. This is because under median regression, the conditional median $M(m_i|x_i)$ is a linear function of x_i, that is,

$M(m_i|x_i) = \beta_{02} + \alpha x_i$. Therefore, $M(m_i|x_i + 1) - M(m_i|x_i) = \beta_{02} + \alpha x_i - \beta_{02} - \alpha(x_i + 1) = \alpha$, and similarly, $M(m_j|x_j + 1) - M(m_j|x_j) = \beta_{02} + \alpha x_j - \beta_{02} - \alpha(x_j + 1) = \alpha$. Note that, other than independence, median regression does not impose any distributional assumptions on e_{2i}, such as homoscedasticity or normality.

To estimate the unknown regression coefficients β_{02} and α in equation 10, median regression takes the approach of the least absolute deviations (LADs), which minimizes the following sum of the absolute deviations,

$$\sum_{i=1}^{n} \left| m_i - \hat{\beta}_{02} - \hat{\alpha} x_i \right|. \quad (11)$$

Compared to the sum of the squared errors, formula 4, the sum of the absolute deviations is less influenced by large errors. Therefore, the LAD estimates are more robust against large errors (e.g., outliers) than the OLS estimates. Computationally, the LAD is more challenging than the OLS because the sum of the absolute deviations, formula 11, is not everywhere differentiable and it cannot be directly minimized by solving the gradient equation. An iterative algorithm is needed to solve the LAD. Specifically, formula 11 can be expressed in terms of an artificial variable u_i as

$$\sum_{i=1}^{n} u_i; \quad (12)$$

subject to the constraints

$$u_i \geq m_i - \beta_{02} - \alpha x_i \quad (13)$$

$$u_i \geq -(m_i - \beta_{02} - \alpha x_i). \quad (14)$$

Noting that minimizing formula 11 is equivalent to minimizing formula 12 under the constraints, the LAD problem can be converted into a standard linear programming problem of optimizing a linear function subject to linear constraints, which can be solved routinely using any available linear programming package. Koenker (2005) provides more technical details on this issue.

Under large sample sizes, median regression estimates follow an asymptotic multivariate normal distribution (Koenker & Bassett, 1978). For example, in formula 11, the median regression estimates $\hat{\theta} = (\hat{\beta}_{02}, \hat{\alpha})$ follow the limiting bivariate normal distribution

$$\sqrt{n}(\hat{\theta} - \boldsymbol{\theta}) \sim N\left(0, \frac{H_n^{-1} J_n H_n^{-1}}{4}\right); \quad (15)$$

where $J_n = n^{-1}\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^T$ and $H_n = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^T f(F^{-1}(0.5))$, with $f(\cdot)$ and $F(\cdot)$ denoting the conditional density and cumulative density function of m_i, respectively. The covariance matrix of $\hat{\theta}$ has the form of a sandwich estimator, which can be estimated using the kernel method proposed by Powell (1991).

Robust Single-level Mediation Analysis

Our robust mediation analysis based on median regression is directly applicable to a general single-level mediation model with multiple mediators and independent variables. For ease of exposition, we focus on the simple mediation model with only one mediator and independent variable (Figure 1). Our approach is based on three median regression equations:

$$y_i = \beta_{01} + \tau x_i + e_{1i} \quad (16)$$

$$m_i = \beta_{02} + \alpha x_i + e_{2i} \quad (17)$$

$$y_i = \beta_{03} + \beta m_i + \tau' x_i + e_{3i}, \quad (18)$$

where the medians of the errors e_{1i}, e_{2i} and e_{3i} equal

To conduct a robust mediation analysis, we fit the median regression equations and estimate the unknown regression parameters using the LAD method. For example, to estimate β_{03}, β and τ', we minimize $\sum_{i=1}^{n} |y_i - \hat{\beta}_{03} - \hat{\beta} m_i - \hat{\tau}' x_i$. After obtaining estimates of regression parameters α and β, the estimate of the mediation effect is given by $\hat{\alpha}\hat{\beta}$.

As mentioned previously, two definitions of the mediation effect, $\tau - \tau'$ and $\alpha\beta$, are equivalent under the standard mediation analysis, in which normality and homoscedasticity are assumed. Although this property does not generally hold in the proposed robust mediation analysis, if errors are independently normally distributed, even without the homoscedasticity assumption, the equivalence of the two definitions also holds for the robust mediation analysis.

Theorem 1. If e_{2i} and e_{3i} are independent and normally distributed, then the equality $\tau - \tau' = \alpha\beta$ corresponds to the same equality in the standard mediation analysis.

Proof: Substituting equation 17 into equation 18, the following equations hold for all values of x_i,

$$y_i = \beta_{03} + \beta(\beta_{02} + \alpha x_i + e_{2i}) + \tau' x_i + e_{3i} = \beta_{03} + \beta\beta_{02} + (\alpha\beta + \tau') x_i + \beta e_{2i} + e_{3i} = \beta_{03} + \beta\beta_{02} + (\alpha\beta + \tau') x_i + e_i, \quad (19)$$

with $e_i = \beta e_{2i} + e_{3i}$. Because e_{2i} and e_{3i} are independently normally distributed, e_i also follows the normal distribution and

$$M(e_i|x_i) = E(\beta e_{2i} + e_{3i}|x_i) = \beta E(e_{2i}|x_i) + E(e_{3i}|x_i) = \beta M(e_{2i}|x_i) + M(e_{3i}|x_i) = 0. \quad (20)$$

That is, the conditional median of e_i is 0. Therefore, by comparing equation 19 with equation 16, we have

$$\tau - \tau' = \alpha\beta. \quad (21)$$

When the condition that e_{2i} and e_{3i} are independent and normally distributed is not satisfied, the value of $\tau - \tau'$ may not equal that of $\alpha\beta$. In this case, one may wonder which measure should be used as the mediation effect. This question is also raised in the standard mediation analysis involving logistic, multinomial or multilevel models, for which $\alpha\beta$ may not equal $\tau - \tau'$. Although subject to debate, we recommend using $\alpha\beta$ as the measure of the mediation effect because it is more in line with the causal interpretation of the mediation effect; that is, $\alpha\beta$ measures how the independent variable causes (or affects) the mediating variable (measured by α), which in turn causes (or affects) the dependent variable (measured by β; MacKinnon, Lockwood, Brown, Wang & Hoffman, 2007; Pearl, 2010). This choice has been adopted by other researchers for multilevel mediation (Bauer, Preacher & Gil, 2006; Kenny, Korchmaros & Bolger, 2003; Krull & MacKinnon, 1999).

We have focused on the simple mediation model with only one mediator and independent variable. Our robust mediation analysis described above is directly applicable to a general single-level mediation model with multiple mediators and independent variables. For instance, assume a two-mediator model (with mediators m_1 and m_2),

$$m_{1i}=\beta_{01}+\alpha_1 x_i+e_{2i} \quad (23)$$

$$m_{2i}=\beta_{02}+\alpha_2 x_i+e_{3i} \quad (24)$$

$$y_i=\beta_{03}+\beta_1 m_{1i}+\beta_2 m_{2i}+\tau' x_i+e_{4i}. \quad (25)$$

To apply the proposed robust method, we first fit median regression models 23, 24 and 25 using the LAD method as previously described and then we can estimate the total mediation effect by $\hat{\alpha}_1\hat{\beta}_1 + \hat{\alpha}_2\hat{\beta}_2$, where $\hat{\alpha}_1$, $\hat{\alpha}_2$, $\hat{\beta}_1$ and $\hat{\beta}_2$ are the estimates of these parameters obtained from the median regression.

Simulation Study of Single-level Mediation Analysis

Simulation description

We conducted two simulations (A and B) to evaluate the performance of the proposed method. Simulation A compares the proposed robust mediation analysis with the standard mediation analysis under various conditions. Specifically, we manipulated three factors: effect size, sample size and error distribution. Following MacKinnon et al. (2002) and Yuan and MacKinnon (2009), we considered three values of α, β and τ' by setting $\alpha = \beta = \tau' =$ 0.14, 0.39 and 0.59, corresponding approximately to small, medium and large effect sizes, respectively, and five values of sample size, n = 50, 100, 200, 500 and 1000. To determine the type I error rates, we also considered the null case of no mediation effect with $\alpha\beta = 0$ by setting $\alpha = 0$ and β =0.39. We focused on four distributions for errors e_{2i} and e_{3i} that are commonly encountered in psychosocial data (Micceri, 1989; Hill & Dixon, 1982). (1) The standard normal distribution $N(0, 1)$. Based on Micceri's survey, only 13.6% of psychometric measures approximately follow normal distributions. (2) A heavy-tailed distribution, i.e., $t_{df=2}$, a t-distribution with 2 degrees of freedom. This distribution has C_{95}=3.61, where C_{95} is a measure of the distribution tail weight and defined as the ratio of the 95th percentile point to the 75th percentile point (Elashoff & Elashoff, 1978). According to Micceri's classification, this heavy-tailed distribution belongs to the double exponential category. Micceri (1989) reported that about 23.2% of the psychometric measures and 60.0% of the criterion mastery measures fall into that category. (3) A contaminated normal distribution, $0.9 \times N(0,1) + 0.1 \times N(0, 10^2)$, i.e., the standard normal distribution $N(0, 1)$ contaminated by $N(0, 10^2)$, a normal distribution with a much larger variance. This distribution has C_{95}=3.07, and thus belongs to the category of extreme contamination based on Micceri's classification. Micceri (1989) reported that about 28.0% of the psychometric measures fall in that category. (4) A normal heteroscedastic distribution, $N(0, x_i^2)$. We will consider skewed error distributions in Simulation B. To simulate data under each parameter setting, we first generated x_i from the standard normal distribution; then conditional on the values of x_i, we simulated m_i and y_i according to equations 17 and 18. Without loss of generality, we assumed that $\beta_{02} = \beta_{03} = 0$ for convenience. We generated a total of 10,000 data sets under each parameter setting.

For each simulated data set, we estimated the mediation effect using the proposed robust mediation analysis and the standard mediation analysis. To compare the performance of the two methods, we calculated the empirical bias, mean squared error (MSE), type I error rate and power for testing the null hypothesis, $\alpha\beta = 0$. Letting $\hat{\alpha}_j\hat{\beta}_j$ denote the estimate of the mediation effect based on the *j*th simulated data set, we obtained the empirical bias and MSE using the following formulae:

$$\text{Empirical bias} = \frac{1}{10000}\sum_{j=1}^{10000}(\hat{\alpha}_j\hat{\beta}_j - \alpha\beta), \quad (26)$$

$$\text{MSE} = \frac{1}{10000}\sum_{j=1}^{10000}(\hat{\alpha}_j\hat{\beta}_j - \alpha\beta)^2. \quad (27)$$

We also calculated the relative root mean squared error (RRMSE), defined as $\text{RRMSE} = \sqrt{\text{MSE}}/\alpha\beta$, as a standardized version of the MSE. The type I error rate (or power) was calculated as the proportion of CIs that do not contain 0 across 10,000 simulations when the mediation effect $\alpha\beta = 0$ (or $\alpha\beta \neq 0$). Both the distribution of the product method and the bias-corrected bootstrap method were used to construct the CI and test the mediation effects. For the standard mediation analysis, two variations of the distribution of the product method were examined: the distribution of the product based on the OLS estimate of variance, equation 7, and the distribution of the product based on the HC3 estimate of variance, equation 9. For the bias-corrected bootstrap method, we used a total of 1,000 bootstrap samples to calculate the bootstrap CI and rejection rate. Recently, Biesanz, Falk and Savalei (2010) showed that the approach of testing a hypothesis based on the CI was not guaranteed to produce accurate type I error rates and power in finite samples, and other methods, in particular hierarchical Bayesian Markov chain Monte Carlo (MCMC) methods, often yielded more accurate results. A comparison of the present approach with the MCMC method will be an interesting topic for future research.

Simulation B was designed to investigate the following question: At what degree of departure from normality does the standard mediation analysis become problematic such that the proposed robust method is preferred? We simulated errors e_{2i} and e_{3i} from a family of distributions with an increasing degree of heavy tails, skewness or heteroscedasticity, and then compared the relative performance of the robust method to that of the OLS method. Specifically, we simulated e_{2i} and e_{3i} in mediation equations 17 and 18 from the Tukey *g*-and-*h* distributions, which are a family of distributions generated by transforming the standard normal variable *z* according to

$$T_{g,h} = \left(\frac{e^{gz} - 1}{g}\right)\exp\left(\frac{hz^2}{2}\right). \quad (28)$$

In the Tukey *g*-and-*h* distribution, the parameter $g \geq 0$ controls asymmetry or skewness, and $h \geq 0$ controls elongation or the extent to which the tails are stretched (relative to the standard normal distribution). Because $(e^{gz} - 1) / g = z + gz^2 / 2! + g^2z^3 / 3! + \cdots$, in the case in which $g = 0$, equation 28 reduces to $T_{g,h} = z\exp(hz^2 / 2)$, a symmetric distribution; and in the case in which $g = h = 0$, $T_{g,h}$ follows a standard normal distribution. A larger value of *g* (or *h*) indicates more severe skewing of the distribution (or heavier tails). To investigate the effects of skewed distributions, we set $h = 0$ and gradually increased the value of *g* from 0 to 1.5 with increments of 0.05, thereby generating a family of skewed distributions with skewness increasing from 0 to 17. To investigate the effects of heavy tails, we set $g = 0$ and gradually increased the value of *h* from 0 to 0.6 with increments of 0.05 to obtain a series of distributions with increasingly heavy tails, with C_{95} ranging from 2.43 to 4.79. Those heavy-tailed distributions stretch the *p*th quantile of the standard normal distribution, say z_p, to $Z_p e^{hz_p^2/2}$. For example, when $h = 0.5$, the resulting heavy-tailed distribution stretches the

90th percentile of the standard normal distribution from 1.28 to 1.93. When $h = 1$, the corresponding distribution provides a close approximation of the Cauchy distribution.

In addition to skewness and heavy tails, the effects of heteroscedasticity were examined by simulating e_{2i} and e_{3i} from a heteroscedastic normal distribution,

$$e_{2i}, e_{3i} \sim N(0, x_i^{\delta}). \quad (29)$$

Under this distribution, the variances of y and m depend on the independent variable x. A large value of δ induces stronger heteroscedasticity, and $\delta = 0$ corresponds to a homoscedastic distribution. We gradually increased the value of δ from 0 to 1.5 with increments of 0.1 to simulate a family of increasingly heteroscedastic distributions. In simulation B, we assumed a sample size of 200 and a median effect size of $\alpha = \beta = \tau' = 0.39$. Under each parameter setting, we simulated 10,000 replications.

Simulation Results

Tables 1 through 3 show the results of simulation A, including the RRMSE (×1,000), type I error rates and power for testing the null hypothesis of no mediation effect under four different error distributions. Both the OLS and proposed robust methods had minimal bias (thus results are not shown), but performed differently in terms of efficiency (i.e., RRMSE; see Table 1). When the error distributions were normal and homoscedastic, the OLS estimate was the maximum likelihood estimate and thus was more efficient than the robust method, as reflected by the smaller RRMSEs. However, as described previously, in practice error distributions are rarely normal and homoscedastic, and instead are often heavy-tailed and/or contaminated. In these cases, the proposed robust estimate can be substantially more efficient than the OLS estimate. For instance, across different effect sizes, the RRMSEs of the robust estimates were about one third (or less than one half) of those for the OLS estimate when the error distributions were heavy-tailed (or contaminated). Comparatively, heteroscedastic errors had less effect on the efficiency, and the RRMSEs were generally comparable between the OLS and robust methods. However, the impact of the heteroscedastic errors on the type I errors was profound, and is described as follows.

For type I errors (see Table 2), albeit with some variation, both the distribution of the product and bootstrap methods generally controlled the type I error rate below the nominal value (i.e., 5%) for OLS, HC3 and robust approaches when the error distributions were normal, heavy-tailed or contaminated. However, when the error distributions were heteroscedastic, the distribution of the product and bootstrap methods led to different performances. Specifically, for the OLS approach, the distribution of the product method based on the OLS estimate of variance led to high type I error rates ranging from 15.9 % to 26.5%; whereas the distribution of the product method based on the HC3 estimate of variance yielded reasonable type I error rates (under 5%). Such differences were due to the use of different variance estimates in these two methods. The distribution of the product method based on the OLS estimate of variance used the OLS estimates of $\sigma^2_{\hat{\alpha}}$ and $\sigma^2_{\hat{\beta}}$, which were biased when the error distributions were heteroscedastic and thus led to high type I error rates. By contrast, the distribution of the product method based on the HC3 estimate of variance utilized the HC3 estimates of $\sigma^2_{\hat{\alpha}}$ and $\sigma^2_{\hat{\beta}}$, which were consistent when the error distributions were heteroscedastic and therefore yielded valid inference. As expected, under the OLS approach, the bootstrap method performed well because it is a nonparametric approach that does not involve estimating $\sigma^2_{\hat{\alpha}}$ and $\sigma^2_{\hat{\beta}}$. As long as the point estimate of the mediation effect is consistent, the bootstrap method is generally valid.

In terms of power, when the error distributions were normal, as expected, the OLS approach was optimal and outperformed the robust approach. However, when the error distributions were heavy-tailed or contaminated, as is often the case in practice, using the proposed robust approach can be substantially more powerful (see Table 3). For example, given a medium effect size and n = 200, and using the bootstrap method, the power of the robust approach was 36% or 55.3% higher than that of the OLS approach when the error distributions were heavy-tailed or contaminated. These results show that although it has been promoted as a desirable method for testing mediation effects in a standard mediation analysis, the bootstrap method cannot entirely address the low efficiency problems of the OLS method when the errors are not normally distributed. Combining the bootstrap method with the proposed robust method was more powerful than using the bootstrap method with the OLS method. The performance of the bootstrap method relative to that of the distribution of the product method differed when used in combination with the OLS or the proposed robust methods. For the robust method, the bootstrap method substantially outperformed the distribution of the product method with substantially higher power (i.e., can be more than 30% higher); whereas for the OLS method, the power of the bootstrap method was only slightly better than the power of the distribution of the product method based on the OLS estimate of variance (i.e., typically <10% improvement) and was comparable to the power of the distribution of the product method based on the HC3 estimate of variance. Note that when the error distributions were heteroscedastic, for the OLS approach, the bootstrap method led to considerably lower power than the distribution of product method, but that is because the latter method had inflated type I error rates.

In summary, simulation A shows that the proposed method is robust to various departures from the standard normality and homoscedasticity assumptions, and yields more efficient estimates and powerful tests than the OLS method in these cases. For the proposed approach, the bootstrap method performs better than the distribution of the product method with relatively higher power. It is worth noting that although the bootstrap method or the distribution of the product method based on the HC3 estimate of variance can improve the performance of the OLS approach, they cannot fully resolve the problems when the errors are not normally distributed. Combining the bootstrap method with the proposed robust method improves the power of mediation analysis.

We now turn to the results of simulation B, which examines the scenario in which the proposed robust method outperforms the standard OLS-based mediation analysis. These results provide some guidance as to which method should be adopted when the assumptions are violated. Figure 2 shows the effects of skewness (i.e., the value of g) on the mean squared error (MSE) and power of the standard and robust mediation analyses. When the skewness (or the value of g) was close to 0, the standard mediation analysis performed slightly better than the robust method with a smaller MSE and higher statistical power. However, when the skewness was greater than 2 (i.e., $g > 0.55$), the proposed robust method outperformed the standard mediation analysis with smaller MSEs. For testing the mediation effect, when the bootstrap method was used, the robust mediation analysis was more powerful than the standard mediation analysis when the skewness was larger than 1.8 (i.e., $g > 0.5$). Micceri (1989) reported that about 18.4% of psychometric measures and 57.1% of criterion mastery measures had levels of skewness greater than 2. In these cases, using the robust method can yield more efficient estimates and more powerful tests.

The performance of the mediation analysis methods under different heavy tails of the distribution is depicted in Figure 3. Compared to the standard mediation analysis, the robust mediation analysis yielded smaller MSEs when $h > 0.15$ (corresponding to $C_{95}>2.89$) and higher power when $h > 0.11$ (corresponding to $C_{95}>2.77$). Micceri (1989) reported that about 51.2% of psychometric measures and 91.4% of criterion mastery measures have extremely heavy tails with C_{95}>2.80. Therefore, in many practical cases, we expect the proposed robust method would perform better than the standard mediation analysis.

As for the effects of the heteroscedastic distributions, Figure 4 suggests that when $\delta > 0.87$, the standard mediation method is inferior to the proposed robust method with larger MSEs. For hypothesis testing, as discussed previously, the main deleterious effect of heteroscedastic distributions is inflated type I error rates; thus in this case we focused on type I error rates rather than power. As shown in Figure 4, for the distribution of the product method based on the OLS estimate of variance, the type I error rate dramatically inflated when $\delta > 0.4$. For instance, when δ=1, the type I error rate of the standard mediation analysis was 18.1% higher than that of the robust mediation analysis when the distribution of the product method based on the OLS estimate of variance was used. In contrast, the distribution of the product method based on the HC3 estimate of variance and bootstrap methods were not sensitive to heteroscedasticity, and differences in the type I error rates between the standard and robust mediation analyses were typically small. This result suggests that in the case of heteroscedastic errors, researchers should use the distribution of the product method based on the HC3 estimate of variance and the bootstrap method.

Example

We illustrate the robust mediation methods using a data set collected from a randomized experiment that investigated the efficacy of a job training intervention program to increase the reemployment rate and enhance the mental health of job seekers (Vinokur & Schul, 1997). A total of 1,801 unemployed workers were randomized to intervention and control groups. Those randomized to the intervention group participated in workshops designed to enhance their job search skills and to provide coping strategies for dealing with setbacks in the job search process. The unemployed workers randomized to the control group received a booklet of job search tips. At the follow-up, the participants' depressive symptoms were measured using the Hopkins Symptom Checklist. One research topic of interest for this study was to investigate whether the effects of the intervention (i.e., independent variable *x*) on the participants' depression symptoms (i.e., dependent variables *y*) were mediated by the participants' job search self-efficacy (i.e, mediating variable *m*). Both the depression symptoms and job search self-efficacy were measured as Likert items and took the form of composite scores. More descriptions of the study can be found in the original article by Vinokur and Schul (1997).

To assess the normality assumption for errors, in Figure 5, we display the normal quantile-quantile plot for the residuals of *y* after regressing on *m* and *x* and the residuals of *m* after regressing on *x*. These plots suggest violations of the normality assumption with skewness and heavy tails. For this randomized study, the residuals are essentially homoscedastic with similar variances between the control and treatment groups.

We applied the proposed robust mediation analysis to this data set and compared the results with those obtained from the standard OLS mediation analysis (Table 4). Under the standard mediation analysis, the 95% CIs based on the bootstrap and distribution of the product methods all included 0, suggesting that the mediation effect was not significant. In contrast, the proposed robust method appeared to be more powerful and detected a (marginally) significant mediation effect: the 95% CIs obtained by the distribution of the product excluded 0, and the 95% CIs based on the bootstrap method marginally excluded 0 with the upper confidence limit equal to zero.

Robust Multilevel Mediation Analysis

Multilevel mediation analysis provides a framework for analyzing mediation effects in correlated data, such as clustered data and longitudinal data. Multilevel mediation analysis is based on multilevel models, which assume that there are at least two levels in the data, an upper level and a lower level (Hox, 2002; Raudenbush & Bryk, 2002). The lower or first-level units (e.g., individuals) are often nested within the upper or second-level units (e.g., groups). Multilevel models account for the correlation among the lower level observations by introducing random effects. An important aspect of multilevel models is that they allow for making inference at the first (lower) level and second (higher) level separately.

Various multilevel mediation analysis methods have been proposed in the literature (Bauer, Preacher & Gil, 2006; Kenny, Kashy & Bolger, 1998; Kenny, Korchmaros & Bolger, 2003; Krull & MacKinnon, 2001; Yuan & MacKinnon, 2009; among others). Almost all of the available methods assume that errors at each level of the model follow normal distributions. However, as pointed out by Kenny et al. (2003), this normality assumption poses a major limitation for multilevel mediation methods. Unlike single-level models, multilevel models are not appropriate for the OLS method because of the correlation among observations. Likelihood-based approaches, e.g., maximum likelihood or Bayesian methods, are typically used to estimate mediation effects in multilevel models. Because the form of the likelihood is directly tied to the assumed distributional assumption, if the normality assumption is violated, the resulting likelihood and the related inference can be invalid. For this reason, violation of the normality assumption is often of more concern for multilevel mediation analysis than for the OLS-based single-level mediation analysis, in which the likelihood is not directly used.

For notational simplicity and clarity, we focus on a two-level mediation model. Let i subscripts refer to the units of the first level (e.g., individuals) and j subscripts refer to the units of the second level (e.g., groups) with $j = 1, \ldots, N$ and $i = 1, \ldots, n_j$. The proposed two-level robust mediation model based on median regression can be expressed as follows for the first level,

$$m_{ij} = \gamma_j + \alpha_j x_{ij} + e_{mij} \quad (30)$$

$$y_{ij} = \delta_j + \beta_j m_{ij} + \tau_j' x_{ij} + e_{yij} \quad (31)$$

where γ_j and δ_j are random intercepts, and α_j, β_j and τ_j' are random slopes. The random intercepts and slopes allow different second-level units, say groups, to have different regression intercepts and slopes. Unlike the majority of the available multilevel mediation analysis methods, which assume that the first-level errors e_{mij} and e_{yij} follow normal distributions, we do not herein impose any distributional assumptions on these errors. We assume only that e_{mij} and e_{yij} are independent and their conditional median is equal to 0, that is, $M(e_{mij}|x_{ij}) = M(e_{yij}|m_{ij}, x_{ij}) = 0$.

The second level of the proposed model describes the distributions of the random effects,

$$\gamma_j = \gamma + u_{\gamma j} \quad (32)$$

$$\alpha_j = \alpha + u_{\alpha j} \quad (33)$$

$$\delta_j = \delta + u_{\delta j} \quad (34)$$

$$\beta_j = \beta + u_{\beta j} \quad (35)$$

$$\tau_j' = \tau' + u_{\tau' j}, \quad (36)$$

where α and β are population (or average) slopes that respectively specify the average effect of the independent variable on the mediating variable, and the average effect of the mediating variable on the dependent variable after controlling the independent variable. The parameters γ and δ are population (or average) intercepts. For multilevel models, depending on the model levels in which the independent, dependent and mediating variable are located, different types of mediation effects can be identified (Kenny, Kashy & Bolger, 1998; MacKinnon, 2008). Here we focus on the population (or average) mediation effect defined by $\alpha\beta$.

For convenience, we assume that the second-level errors $u_{\gamma j}$, $u_{\alpha j}$, $u_{\delta j}$, $u_{\beta j}$, and $u_{\tau' j}$ are independent and follow normal distributions with a mean of zero and unique variances $\sigma^2_{\gamma_j}, \sigma^2_{\alpha_j}, \sigma^2_{\delta_j}, \sigma^2_{\beta_j}$, and $\sigma^2_{\tau'_j}$, respectively. However, this second-level normality assumption is not critical. Verbeke and Lesaffre (1997) showed that even when the normality assumption of the second-level errors is violated, the maximum likelihood estimates of α, β and τ' remain consistent.

The estimation of multilevel median regression models is substantially more complicated than that of the single-level median regression model. We illustrate the estimation procedure using the *m*-equations, i.e., equations 30, 32 and 33, which specify the effect of the independent variable *x* on the mediating variable *m*. To do that, it is convenient to substitute equations 32 and 33 into equation 30 and re-express the *m*-equations as follows,

$$m_{ij} = \gamma + \alpha x_{ij} + u_{\gamma j} + u_{\alpha j} x_{ij} + e_{mij} \quad (37)$$

$$u_{\gamma j} \sim N\left(0, \sigma^2_{\gamma_j}\right) \quad (38)$$

$$u_{\alpha j} \sim N(0, \sigma^2_{\alpha_j}). \quad (39)$$

The unknown regression coefficients in the *m*-equations are estimated by minimizing the following l_2 penalized absolute deviations (Geraci & Bottai, 2007; Yuan & Yin, 2010),

$$\sum_{j=1}^{N} \sum_{i=1}^{n_j} |m_{ij} - \gamma - \alpha x_{ij} - u_{\gamma j} - u_{\alpha j} x_{ij}| + u^2_{\gamma j}/2\sigma^2_{\gamma_j} + u^2_{\alpha j}/2\sigma^2_{\alpha_j}, \quad (40)$$

in which the first term is the sum of the absolute deviations coming from median regression equation 37, and the second and third terms are the l_2 penalty induced by the random effects $u_{\gamma j}$ and $u_{\alpha j}$ in equations 38 and 39.

Before describing a method to minimize the l_2 penalized absolute deviations, we introduce a relationship between the absolute deviation and the Laplace distribution (also known as a double exponential distribution). This relationship will be utilized to maximize the l_2 penalized absolute deviations and obtain the estimates of the unknown parameters. The

Laplace distribution has a density function $f(m|\theta,\lambda)=\frac{1}{2\lambda}\exp(-|m-\theta|/\lambda)$ where θ and λ are location and scale parameters, respectively. For our purpose, we focus on the Laplace distribution with a scale parameter of 1, that is,

$$f(m|\theta)=1/2\exp(-|m-\theta|). \quad (41)$$

Setting $\theta = \gamma + \alpha x_{ij} + u_{\gamma j} + u_{\alpha j}x_{ij}$, the negative logarithm of the Laplace density function is

$$-\log\left(f(m|\theta)\right)=|m-\gamma-\alpha x_{ij}-u_{\gamma j}-u_{\alpha j}x_{ij}|+\log(2). \quad (42)$$

which is equivalent to the first term of formula 40, that is, the absolute deviations. We call two functions equivalent if they differ by only a constant. For example, equation 42 and the first term of formula 40 are equivalent because they differ only by a constant of log(2). Given this relationship, and also noting that the last two terms in formula 40 are equivalent to the negative logarithm of the standard normal distribution random effects $u_{\gamma j}$ and $u_{\alpha j}$, it follows that the l_2 penalized absolute deviations are equivalent to the negative log likelihood of the following two-level "working" model (Yuan & Yin, 2010)

$$m_{ij}\sim LD(\gamma+\alpha x_{ij}+u_{\gamma j}+u_{\alpha j}x_{ij})\ u_{\gamma j}\sim N\left(0,\sigma^2_{\gamma_j}\right)\ u_{\alpha j}\sim N(0,\sigma^2_{\alpha_j}), \quad (43)$$

where $LD(\theta)$ denotes a Laplace distribution with a location parameter of θ. Therefore minimizing the l_2 penalized absolute deviations, formula 40, is equivalent to maximizing the likelihood function of working model 43. In other words, by utilizing the relationship between the absolute deviation and the Laplace distribution, we cast the problem of minimizing the l_2 penalized absolute deviations into a familiar problem of maximizing the likelihood of working model 43. Note that model 43 is just a "working" model; the distributional assumption we assumed on m_{ij} is completely artificial and is used solely to match the l_2 penalized absolute deviations. The model we actually assume is depicted in the *m*-equations 30, 32 and 33.

Along the same lines, the *y*-equations, i.e., equations 31, 34, 35 and 36, are estimated by minimizing the l_2 penalized absolute deviations

$$\sum_{j=1}^{n}\sum_{i=1}^{n_j}|y_{ij}-\gamma-\beta m_{ij}-\tau' x_{ij}|+\gamma_j^2/2\sigma^2_{\gamma_j}+\beta_j^2/2\sigma^2_{\beta_j}+\tau'^2_j/2\sigma^2_{\tau'_j}. \quad (44)$$

which can be cast as the problem of fitting the following two-level working model,

$$y_{ij}\sim LD(\gamma_y-\beta m_{ij}-\tau' x_{ij}-u_{\gamma j}-u_{\beta j}m_{ij}-u_{\tau' j}x_{ij})\ u_{\gamma j}\sim N\left(0,\sigma^2_{\gamma_j}\right)\ u_{\beta j}\sim N(0,\sigma^2_{\beta_j})\ u_{\tau' j}\sim N(0,\sigma^2_{\tau'_j}). \quad (45)$$

One convenient way to fit these working models is the Bayesian approach (Yuan & Yin, 2010). As discussed by Yuan and MacKinnon (2009), the Bayesian approach is particularly appealing for complex multilevel mediation analysis because once the model is fitted, the inference about the mediation effect is straightforward and exact without relying on large-sample approximation. Specifically, to fit working model 43, we assign independent noninformative flat priors to γ and α of the form $\gamma \infty 1$ and $\alpha \infty 1$, and independent vague inverse gamma priors IG(10^{-6}, 10^{-6}) to variances $\sigma^2_{\gamma_j}$ and $\sigma^2_{\alpha_j}$. We then fit the model using a Gibbs sampler, an iterative algorithm that simulates the posterior distributions of unknown parameters by sequentially sampling unknown parameters from their conditional distributions; that is, we sample each parameter from the distribution of that parameter conditioned on the data and all other parameters. As the output of the Gibbs sampler, posterior samples of α, say $\{\alpha^{(t)}, t = 1, \ldots, T\}$, are obtained from T iterations. Similarly, the posterior sample of β, say $\{\beta^{(t)}, t = 1, \ldots, T\}$, is estimated by fitting working model 45 through the Gibbs sampler with independent noninformative flat priors for γ, β, and τ′ (i.e., $\gamma \propto 1$, $\beta \propto 1$, and $\tau' \propto 1$) and independent vague inverse gamma priors IG(10^{-6}, 10^{-6}) for variances $\sigma^2_{\gamma_j}$, $\sigma^2_{\beta_j}$ and $\sigma^2_{\tau'_j}$.

The inference of the population mediation effect αβ is based on the posterior samples of αβ, which can be easily obtained by multiplying the posterior samples of α with the posterior samples of β, that is, $\{\alpha^{(t)}\beta^{(t)}, t = 1, \ldots, T\}$. Then an estimate of the population mediation effect αβ is the posterior mean of αβ, given by

$$\widehat{\alpha\beta} = \sum_{t=1}^{T} \alpha^{(t)}\beta^{(t)}/T. \quad (46)$$

and the 95% credible interval of αβ is given by ($q_{0.025}$, $q_{0.975}$), where $q_{0.025}$ and $q_{0.975}$ are 0.025 and 0.975 sample percentiles of the posterior samples of αβ.

Simulation Study of Multilevel Mediation Analysis

Simulation description

We conducted a simulation study to evaluate the performance of the proposed robust multilevel mediation analysis. Loosely following the simulation setting of Bauer et al. (2006), we set γ = δ = 0, $\sigma^2_{\gamma_j}=0.6, \sigma^2_{\delta_j}=0.4, \tau'=0.2, \sigma^2_{\tau'_j}=0.04$ and $\sigma^2_{\alpha_j}=\sigma^2_{\beta_j}=0.16$. We manipulated three design factors: (1) the effect size of the average indirect effect; (2) the sample size; and (3) the distributions of the level 1 errors. For the first factor, we set α = β = 0.3 or α = β = 0.6 to represent a small and large average indirect effect. To investigate the type I error rate, we also considered the null case of no mediation effect (i.e., αβ = 0) with α = 0 and β = 0.3. For the second factor, we set the number of level-two units to N = 25, 50, 100 or 200, and set the number of observations per level-two unit to n_j = 4, 8 or 16. For the third factor, the distributions of the level-one errors e_{yij} and e_{mij}, we considered four different distributions: (a) the standard normal distribution, $N(0, 1)$; (b) a heavy-tailed t-distribution, $t_{df=2}$; (c) a contaminated normal distribution, $0.9 \times N(0, 1) + 0.1 \times N(0, 10^2)$; and (d) a normal heteroscedastic distribution, $N(0, x_i^2)$. Together, these three design factors yielded 192 scenarios in a factorial design. Under each scenario, we simulated 1,000 samples of data. We compared the proposed method with the standard maximum likelihood approach. For the latter approach, we fitted the multilevel model using SAS PROC MIXED and used the variance formula derived by Bauer et al. (2006) to construct the CI. Under each simulation scenario, we calculated the empirical bias (×1,000), RRMSE (×1,000), coverage rate of the 95% CI, and rejection rate based on 1,000 replications.

Simulation results

Across different error distributions, the MLE and proposed robust estimates both had minimal bias (results not shown), but displayed different levels of efficiency in terms of RRMSE (Table 5). Because the RRMSEs were very similar for small and large effect sizes, Table 5 shows the results for only the small effect size. When the error distributions were normal and homoscedastic, the MLE estimates were optimal and outperformed the robust estimates. The loss of efficiency from using the robust estimates was minor, and the RRMSEs were roughly comparable between the MLE and robust estimates. However, if the normality assumptions were not satisfied, the gain from using the proposed robust method could be substantial. Specifically, when the error distributions were heavy-tailed, the

RRMSE of the robust estimates was often less than one third of that of the MLE; and when the error distributions were contaminated, the RRMSE of the robust estimates was typically less than one half of that of the MLE. The MLE estimate was not sensitive to the heteroscedasticity and yielded RRMSEs comparable to those of the robust estimate when the error distributions were heteroscedastic.

In terms of hypothesis testing, the MLE and proposed robust methods were able to control the type I error rates below the nominal value of 5% (results not shown), but possessed different power (see Table 6). When the error distributions were normal, as expected, the MLE was the most efficient method and led to higher power than the robust method. However, when the error distributions were heavy-tailed, the robust method could be substantially more powerful than the MLE. For example, when N=100 and n_j=4, the power of the robust method was 80% while that for the MLE was only 1.8%. Similarly, when the error distributions were contaminated, the power of the robust method was often 5 to 10 times higher than that of the MLE. When the error distributions were heteroscedastic, the performance of the two methods was comparable, suggesting that heteroscedastic errors may not be of particular concern for multilevel models.

Discussion

We have proposed a robust mediation analysis method based on median regression. Unlike the standard OLS or maximum likelihood approaches, the median regression minimizes the absolute deviations and thus is robust to the violation of homoscedastic and normal assumptions. The simulation studies show that the proposed mediation analysis is robust to various departures from the standard normal and homoscedastic assumptions, including heavy-tailed, skewed, outlier contaminated, and heteroscedastic error distributions. In these cases, the resulting estimates are more efficient than the standard OLS estimates and yield better statistical power to test mediation effects. We further extended the robust method to multilevel models to accommodate correlated data. We used a Bayesian approach to estimate the proposed multilevel mediation model. The simulation study demonstrated that the proposed method outperformed the standard maximum likelihood-based approach and resulted in a smaller MSE and higher statistical power.

Another interesting finding of this study is that the widely advocated bootstrap method cannot fully address the problems of the standard OLS method when the error distributions are heavy-tailed, skewed, or outlier contaminated. Although in such cases the bootstrap method performs better than other methods, such as the distribution of the product method, its performance is not satisfactory. Combining the bootstrap method with the proposed robust method can substantially improve the power of testing the mediation effect. As shown in the simulation study, in some cases, this approach doubles the power obtained by using the bootstrap method with the OLS method.

Compared to the standard OLS method, making estimations using the proposed robust method is more involved and needs an iterative fitting procedure. Fortunately, many available statistical software packages provide functions to efficiently implement median regression, which substantially facilitates the application of the proposed robust mediation analysis. We have implemented the proposed robust mediation analysis for a single-level model using a SAS macro (SAS 9.1, SAS Institute, 2003) and R (an open source statistical computing and graphics software; R Development Core Team, 2008), which are available for free downloading from the authors' website. The SAS macro is provided in the Appendix.

The standard OLS or maximum likelihood approaches are optimal when the assumptions of heteroscedasticity and normality hold, whereas the proposed robust method performs better when these assumptions are violated. In order to determine which procedure to use (the OLS or robust methods), researchers should examine the extent to which the assumptions have been violated. As a rule of thumb, the robust method should be seriously considered under the following scenarios: if there are outliers, if the skewness of the error distribution is larger than 2, or if heavy tails exist for which the 90^{th} percentile of the standardized errors is larger than 1.45. Barnett and Lewis (1994) provided various methods to identify and test whether a data point is an outlier. However, we want to emphasize that this rule of thumb is just a rough guideline, and our intention here is not to develop a new method to replace the standard OLS and maximum likelihood methods, but to propose a (robust) method that is supplementary to the standard methods. When violations of normality and homoscedasticity assumptions are detected, instead of choosing a single method to use, we prefer to apply both the standard and robust methods and then compare their results as a form of sensitivity analysis. That is, if the results from the two methods are consistent (e.g., both methods suggest significant or insignificant mediation effects), we obtain more confidence in the standard method; and if the results from the two methods differ (e.g., one method yields significant results and the other yields insignificant results), we then should acknowledge potentially large uncertainties inherent in the analysis results, and probably use simulation studies to assess which method may be more reliable for a specific application setting.

In the multilevel mediation model, we assume that the second-level errors are independent and follow normal distributions. Although this normality assumption is far less restrictive than that of the first-level error distributions in the standard mediation analysis (Verbeke & Lesaffre, 1997), relaxing this distributional assumption may further improve the robustness of the proposed method, especially when the sample size is small or moderate. One possible approach is to model the distributions of the second-level errors as a mixture of normal distributions in order to handle possible deviations from the normal distribution. We can also relax the independence assumption by allowing for correlations among the second-level errors (or random effects). However, although theoretically appealing, such an extension can be problematic in practice because the observed data typically contain little information regarding the correlations among the random effects, and consequently the resulting estimates can be rather unstable or even not estimable (Bauer et al., 2006). In practice, the assumption of independence among random effects is often reasonable and useful to improve the estimation stability of the model.

This paper focused on a new method to assess mediation that is more accurate for the actual distribution of variables obtained in research studies. In the application of this new method as well as other mediation methods, it is important to keep in mind that the mediation model is a causal model and is longitudinal in that x causes m and m causes y. Statistical analysis of any three variables by itself does not provide evidence confirming true mediation or other third-variable effects like confounding (MacKinnon, Krull, & Lockwood, 2000). More information on recent developments in causal inference for mediation models can be obtained elsewhere (Imai, Keele, & Tingley, 2010; Jo et al., 2011; Muthen, 2011; Pearl, 2010; VanderWeele, 2010; West, 2011). Assumptions and prior knowledge of empirical and theoretical relations are needed to justify the sequence x to m to y. For example, if x represents a random assignment, then x comes before m and y. The precedence of m to y compared to y to m is based on theory and prior research. We agree that ideally the change in m is used to predict the change in y in a longitudinal study. As for many developments in mediation analysis, we feel it is clearer to start with the simplest mediation model, with one mediator to develop and describe the results from the first principles. The most reasonable approach is to investigate mediation within a program of research with information from a variety of research studies and designs, including qualitative studies and clinical judgment.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

A program of research is how mediating variables have been identified in past research, e.g., cognitive dissonance theory in psychology and the genetic theory for how the characteristics of the parents are related to those of the offspring. Mediation analysis based on median regression is an important tool in this endeavor because it can most easily handle distributional problems and, most important for applied researchers, can handle outlier observations in a reasonable way.

Acknowledgments

This work was partially supported by the National Cancer Institute grant R01 CA154591 awarded to Ying Yuan and Health Service Grants DA09757 and MH40859 awarded to David MacKinnon.

References

Baron RM, Kenny DA. The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of Personality and Social Psychology. 1986; 51:1173–1182. [PubMed: 3806354]

Barnett, V.; Lewis, T. Outliers in Statistical Data. Third Edition. Chichester: Wiley; 1994.

Bauer DJ, Preacher KJ, Gil KM. Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: New procedures and recommendations. Psychological Methods. 2006; 11:142–163. [PubMed: 16784335]

Biesanz JC, Falk CF, Savalei V. Assessing mediational models: testing and interval estimation for indirect effects. Multivariate Behavioral Research. 2010; 45:661–701.

Bollen KA, Stine RA. Direct and indirect effects: classical and bootstrap estimates of variability. Sociological Methodology. 1990; 20:115–140.

Bryk AS, Raudenbush SW. Heterogeneity of variance in experimental studies: A challenge to standard interpretations. Psychological Bulletin. 1988; 104:396–404.

Collins LM, Graham JW, Flaherty BP. An alternative framework for defining mediation. Multivariate Behavioral Research. 1998; 33:295–312.

Efron B. Bootstrap methods: another look at the jackknife. The Annals of Statistics. 1979; 7:1–26.

Efron B. Better bootstrap confidence intervals. Journal of the American Statistical Association. 1987; 82:171–185.

Elashoff, JD.; Elashoff, RM. Effects of errors in statistical assumptions. In: Kruskal, WH.; Tanur, JM., editors. International Encyclopedia of Statistics. New York: Free Press; 1978. p. 229-250.

Erceg-Hurn DM, Mirosevich VM. Modern robust statistical methods. American Psychologist. 2008; 63:591–601. [PubMed: 18855490]

Finch JF, West SG, MacKinnon DP. Effects of sample size and nonnormality on the estimation of mediated effects in latent variable models. Structural Equation Modeling. 1997; 4:87–107.

Geraci M, Bottai M. Quantile regression for longitudinal data using the asymmetric Laplace distribution. Biostatistics. 2007; 8:140–154. [PubMed: 16636139]

Grissom RJ. Heterogeneity of variance in clinical data. Journal of Consulting and Clinical Psychology. 2000; 68:155–165. [PubMed: 10710850]

Grissom RJ, Kim JJ. Review of assumptions and problems in the appropriate conceptualization of effect size. Psychological Methods. 2001; 6:135–146. [PubMed: 11411438]

Grissom, RJ.; Kim, JJ. Effect Sizes for Research: A Broad Practical Approach. Mahwah, NJ: Erlbaum; 2005.

Hayes AF. PROCESS: A versatile computational tool for observed variable mediation, moderation, and conditional process modeling [White paper]. 2012 Retrieved from http://www.afhayes.com/public/process2012.pdf.

Hayes AF, Cai L. Using heteroscedasticity-consistent standard error estimators in OLS regression: An introduction and software implementation. Behavior Research Methods. 2007; 39:709–722. [PubMed: 18183883]

Hill M, Dixon WJ. Robustness in real life: A study of clinical laboratory data. Biometrics. 1982; 38:377–396. [PubMed: 7115869]

Hogg RV. Adaptive robust procedures: A partial review and some suggestions for future applications and theory. American Statistical Association Journal. 1974; 69:909–926.

Hox, JJ. Multilevel Analysis: Techniques and Applications. Mahwah, NJ: Erlbaum; 2002.

Imai K, Keele L, Tingley D. A general approach to causal mediation analysis. Psychological Methods. 2010; 15:309–334. [PubMed: 20954780]

Jo B, Stuart EA, MacKinnon DP, Vinokur AD. The use of propensity scores in mediation analysis. Multivariate Behavioral Research. 2011; 46:425–452. [PubMed: 22399826]

Kenny, DA.; Kashy, DA.; Bolger, N. Data analysis in social psychology. In: Gilbert, D.; Fiske, S.; Lindzey, G., editors. Handbook of Social Psychology. New York: McGraw-Hill; 1998.

Kenny DA, Korchmaros JD, Bolger N. Lower level mediation in multilevel models. Psychological Methods. 2003; 8:115–128. [PubMed: 12924810]

Keppel, G.; Wickens, TD. Design and Analysis: A Researcher's Handbook. 4th ed.. Upper Saddle River, NJ: Pearson; 2004.

Keselman HJ, Algina J, Lix LM, Wilcox RR, Deering K. A generally robust approach for testing hypotheses and setting confidence intervals for effect sizes. Psychological Methods. 2008; 13:110–129. [PubMed: 18557681]

Keselman HJ, Huberty CJ, Lix LM, Olejnik S, Cribbie RA, Donahue B, et al. Statistical practices of educational researchers: An analysis of their ANOVA, MANOVA, and ANCOVA analyses. Review of Educational Research. 1998; 68:350–386.

Koenker, R. Quantile Regression. Cambridge: Cambridge University Press; 2005.

Koenker R, Bassett G. Regression quantiles. Econometrica. 1978; 46:33–50.

Kraemer HC, Wilson GT, Fairburn CG, Agras WS. Mediators and moderators of treatment effects in RCTs. Archives of General Psychiatry. 2002; 59:877–883. [PubMed: 12365874]

Krull JL, MacKinnon DP. Multilevel mediation modeling in group-based intervention studies. Evaluation Review. 1999; 23:418–444. [PubMed: 10558394]

Krull JL, MacKinnon DP. Multilevel modeling of individual and group level mediated effects. Multivariate Behavioral Research. 2001; 36:249–277.

Lix LM, Keselman JC, Keselman HJ. Consequences of assumption violations revisited: A quantitative review of alternatives to the one-way analysis of variance "F" test. Review of Educational Research. 1996; 66:579–619.

Lockwood, CM. Effects of nonnormality on tests of mediation. Unpublished master's thesis. Tempe, AZ: Arizona State University; 1999.

Lockwood, CM.; MacKinnon, DP. Proceedings of the Twenty-Third Annual SAS Users Group International Conference. Cary, NC: SAS Institute; 1998. Bootstrapping the standard error of the mediated effect; p. 997-1002.

Lomnicki ZA. On the distribution of products of random variables. Journal of the Royal Statistical Society, Series B. 1967; 29:513–524.

Long JS, Ervin LH. Using heteroscedasticity consistent standard errors in the linear regression model. The American Statistician. 2000; 54:217–224.

MacKinnon, DP. Introduction to Statistical Mediation Analysis. New York, NY: Erlbaum; 2008.

MacKinnon DP, Dwyer JH. Estimating mediating effects in prevention studies. Evaluation Review. 1993; 17:144–158.

MacKinnon DP, Krull JL, Lockwood CM. Equivalence of the mediation, confounding, and suppression effect. Prevention Science. 2000; 1:173–181. [PubMed: 11523746]

MacKinnon DP, Lockwood CM, Brown CH, Wang W, Hoffman JM. The intermediate endpoint effect in logistic and probit regression. Clinical Trials. 2007; 4:499–513. [PubMed: 17942466]

MacKinnon DP, Lockwood CM, Hoffman JM, West SG, Sheets V. A comparison of methods to test mediation and other intervening variable effects. Psychological Methods. 2002; 7:83–104. [PubMed: 11928892]

MacKinnon DP, Lockwood CM, Williams J. Confidence limits for the indirect effect: Distribution of the product and resampling methods. Multivariate Behavioral Research. 2004; 39:99–128. [PubMed: 20157642]

MacKinnon DP, Warsi G, Dwyer JH. A simulation study of mediated effect measures. Multivariate Behavioral Research. 1995; 30:41–62. [PubMed: 20157641]

Meeker, WQ.; Cornwell, LW.; Aroian, LA. Selected Tables in Mathematical Statistics, Volume VII: The Product of Two Normally Distributed Random Variables. Providence, RI: American Mathematical Society; 1981.

Micceri T. The unicorn, the normal curve, and other improbable creatures. Psychological Bulletin. 1989; 105:156–166.

Muthen B. Applications of causally defined direct and indirect effects in mediation analysis using SEM in Mplus. 2011 Unpublished manuscript.

Pearl, J. Technical Report. Los Angeles, CA: University of California, Los Angeles; 2010. The mediation formula: A guide to the assessment of causal pathways in non-linear models.

Powell, JL. Estimation of monotonic regression models under quantile restrictions. In: Barnett, W.; Powell, J.; Tauchen, G., editors. Nonparametric and Semiparametric Methods in Econometrics. Cambridge: Cambridge University Press; 1991.

R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2008.

Raudenbush, SW.; Bryk, AS. Hierarchical Linear Models: Applications and Data Analysis Methods. 2nd ed.. Newbury Park, CA: Sage; 2002.

Rubin D. Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of Educational Psychology. 1974; 66:688–701.

SAS Institute. SAS Version 9.1. Cary, NC: 2003.

Schrader RM, Hettmansperger TP. Robust analysis of variance. Biometrika. 1980; 67:93–101.

Serlin RC, Harwell MR. More powerful tests of predictor subsets in regression analysis under nonnormality. Psychological Methods. 2004; 9:492–509. [PubMed: 15598101]

Shrout PE, Bolger N. Mediation in experimental and nonexperimental studies: New procedures and recommendations. Psychological Methods. 2002; 7:422–445. [PubMed: 12530702]

Tomarken AJ, Serlin RC. Comparison of ANOVA alternatives under variance heterogeneity and specific noncentrality structures. Psychological Bulletin. 1986; 99:90–99.

VanderWeele TJ. Bias formulas for sensitivity analysis for direct and indirect effects. Epidemiology. 2010; 21:540–551. [PubMed: 20479643]

VanderWeele TJ, Vansteelandt S. Conceptual issues concerning mediation, interventions and composition. Statistics and Its Interface. 2009; 2:457–468.

Verbeke G, Lesaffre E. The effect of misspecifying the random-effects distribution in linear mixed models for longitudinal data. Computational Statistics & Data Analysis. 1997; 23:541–556.

Vinokur A, Schul Y. Mastery and inoculation against setbacks as active ingredients in the jobs intervention for the unemployed. Journal of Consulting and Clinical Psychology. 1997; 65:867–877. [PubMed: 9337505]

Wegman EJ, Carroll RJ. A Monte Carlo study of robust estimators of location. Communications in Statistics, A6. 1977:795–812.

West SG. Editorial: Introduction to the special section on causal inference in cross sectional and longitudinal mediational models. Multivariate Behavioral Research. 2011; 46:812–815.

Wilcox RR. How many discoveries have been lost by ignoring modern statistical methods? American Psychologist. 1998; 53:300–314.

Wilcox RR, Keselman HJ. Modern robust data analysis methods: Measures of central tendency. Psychological Methods. 2003; 8:254–274. [PubMed: 14596490]

Yuan Y, MacKinnon DP. Bayesian mediation analysis. Psychological Methods. 2009; 14:301–322. [PubMed: 19968395]

Yuan Y, Yin G. Bayesian quantile regression for longitudinal studies with nonignorable missing data. Biometrics. 2010; 66:105–114. [PubMed: 19459836]

Zu J, Yuan K. Local influence and robust procedures for mediation analysis. Multivariate Behavioral Research. 2010; 45:1–44.

Appendix

Herein we provide the SAS macro for conducting the proposed robust mediation analysis for the single mediation model. This macro utilizes a SAS procedure called lav, which is available in the standard SAS system, to conduct median regression. The percentile and bias-corrected bootstrap methods are used to construct the confidence interval.

To use the macro, data should be organized in the standard format: columns are variables of interest (say *y*, *m* and *x*), rows are study units, and "." denotes missing values. For input parameters, *y* is the dependent variable; *m* is the mediating variable, *x* is the independent variable; *dataset* is the name of the dataset containing the above variables, *nboot* is the number of replicates used to calculate the bootstrap confidence interval; and *alphalev* is the type I error. As the output of the macro, A and B are the estimates of α and β; AB is the estimate of mediation effect αβ, ab_lb and ab_ub are the lower and upper bounds of the 95% confidence interval when alphalev=.05 (the default value). Assuming that the data set is named meddata, the robust mediation macro is called with the statement

```
%robustmed(y, m, x, meddata, nboot=1000, alphalev=.05);
```

Below is the SAS macro for conducting the robust mediation analysis:

```
%macro robustmed(y, m, x, dataset, nboot=1000, alphalev = .05);
/* robust mediation analysis for the sample*/
proc iml ;
use &dataset;
read all;
one = J(nrow(&x),1);
mcovariate = one ‖ &x ;
ycovariate = one ‖ &m ‖ &x;
opt= {.0. 1};
call lav(rc1, xr1, mcovariate, &m,,opt); /* median regression of m on x */
call lav(rc2, xr2, ycovariate, &y,,opt); /* median regression of y on m and
x */
ab = xr1[2]||xr2[2]||xr1[2]*xr2[2];
CREATE abest from ab[COLNAME={a b ab}];
APPEND from ab;
quit;
data _NULL_;
set abest;
call symput('sampest', ab);
run;
/* calculate bootstrap confidence interval */
/* generate bootstrap samples */
ods listing close;
proc surveyselect data= &dataset out=bootsample seed = 1347 method = urs
samprate = 1 outhits rep =
```

```
&nboot;
run;
/* fit bootstrap samples */
proc iml;
use bootsample;
read all;
one = J(nrow(&x),1);
mcovariate = one ‖ &x ;
ycovariate = one ‖ &m ‖ &x;
opt= {.0. 1};
n = nrow(&x)/&nboot;
r1 = 1;
r2 = n;
abboot = j(&nboot, 1, -99);
do rep=1 to &nboot;
call lav(rc1, xr1, mcovariate[r1:r2,], &m[r1:r2,],,opt); /* median
regression of m on x */
call lav(rc2, xr2, ycovariate[r1:r2,], &y[r1:r2,],,opt); /* median
regression of y on m and x */abboot[rep, 1] = xr1[2]*xr2[2];
r1 = r1 + n;
r2 = r2 + n;
end;
CREATE estimates from abboot[COLNAME={ab}];
APPEND from abboot;
quit;
/* calculate confidence interval using the percentile bootstrap method*/
%let lpct = %sysevalf(&alphalev/2*100);
%let upct = %sysevalf((1 - &alphalev/2)*100);
proc univariate data = estimates;
var ab;
output out=pboot mean = r2hat pctlpts=&lpct &upct pctlpre = ab pctlname =
_lb _ub ;
run;
/* calculate confidence interval using the bias-corrected bootstrap method*/
/* Get p and z0 for bias-corrected bootstrap */
data findp; set estimates;
if ab > &sampest then p=1; else p=0;
run;
proc means data=findp noprint;
output out=findp2 mean(p)=p;
run;
data _NULL_; set findp2;
z0=probit(1-p);
call symput('z0',z0);
run;
/* Find percentiles to get for bias-corrected bootstrap */
data _NULL_;
zp=probit(&upct/100);
roundpoint=100/&nboot;
bcbzlo=(2*&z0)-zp;
```

```
bcbzup=(2*&z0)+zp;
bcbplo=probnorm(bcbzlo);
bcbpup=probnorm(bcbzup);
bcbpctlo=round(bcbplo*100,roundpoint);
bcbpctup=round(bcbpup*100,roundpoint);
call symput('bcbpctlo',bcbpctlo);
call symput('bcbpctup',bcbpctup);
run;
/* get confidence interval */
proc univariate data = estimates;
var ab;
output out=bcboot pctlpts=&bcbpctlo &bcbpctup pctlpre= ab pctlname=_lb _ub;
run;
/* print results */
data output1;
merge abest pboot;
method = "percentile bootstrap ";
run;
data output2;
merge abest bcboot;
method = "bias-corrected bootstrap";
run;
data output;
set output1 output2;
label ab_lb="lower &lpct% limit" ab_ub="upper &upct% limit";
run;
options label;
run;
ods listing;
proc print data = output label;
var method a b ab ab_lb ab_ub;
run;
%mend;
```

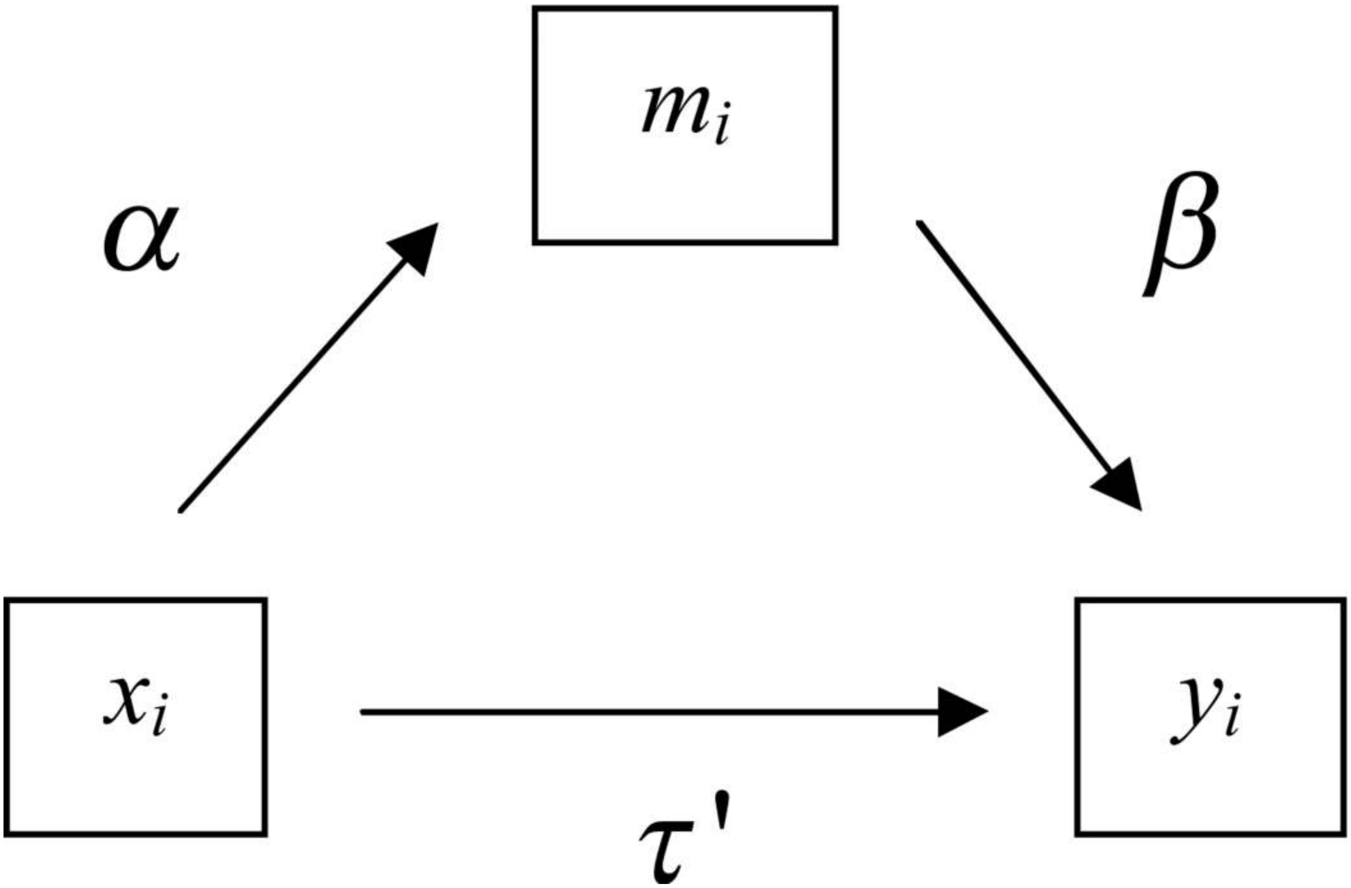

Figure 1.
Path diagram for the single-level mediation model.

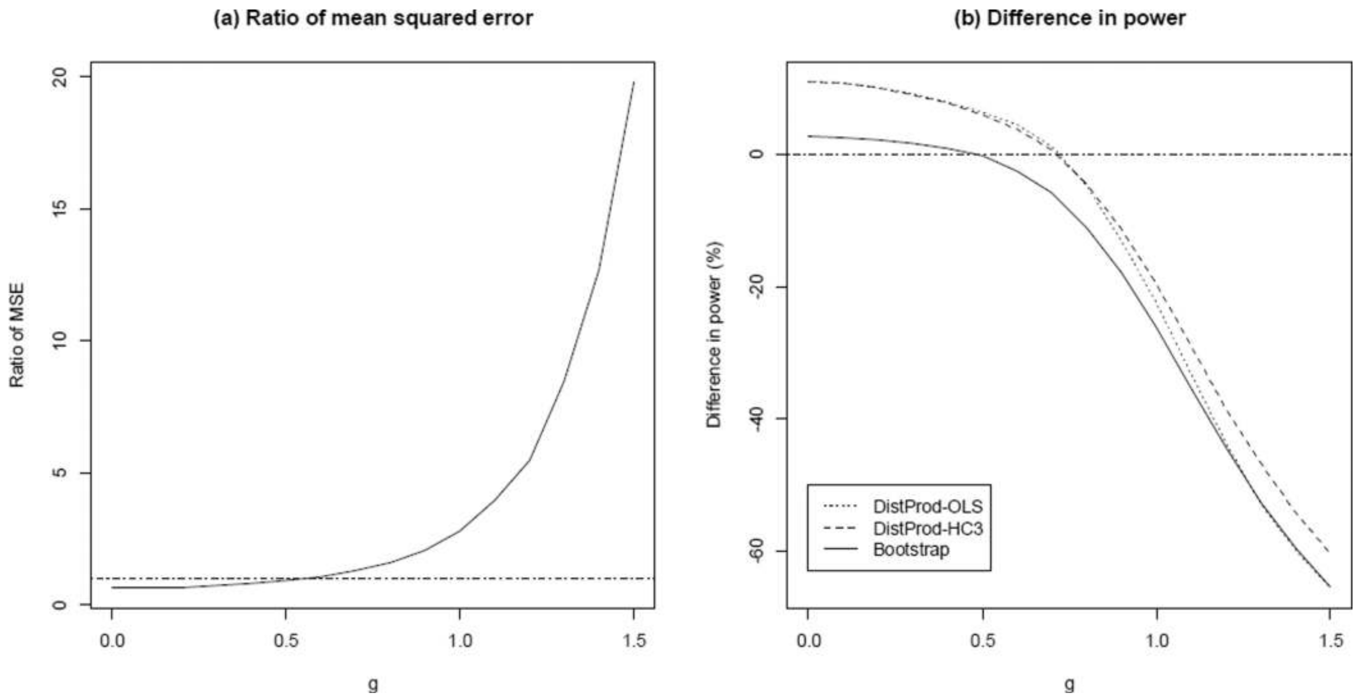

Figure 2.

Performance of standard mediation analysis vs. proposed robust mediation analysis when data are generated from the Tukey *g*-and-*h* family of distributions with different values of *g*. A larger value of *g* represents a higher degree of skewness. Panel (a) shows the ratio of the mean squared error (MSE). For reference, a ratio of 1 is indicated by the dashed horizontal line. Panel (b) shows the difference in power. DistProd-OLS and DistProd-HC3 denote the distributions of the product methods based on the ordinary least squares (OLS) and heteroscedasticity consistent covariance (HC3) estimates of variance, respectively. A difference of 0 is indicated by the dashed horizontal line.

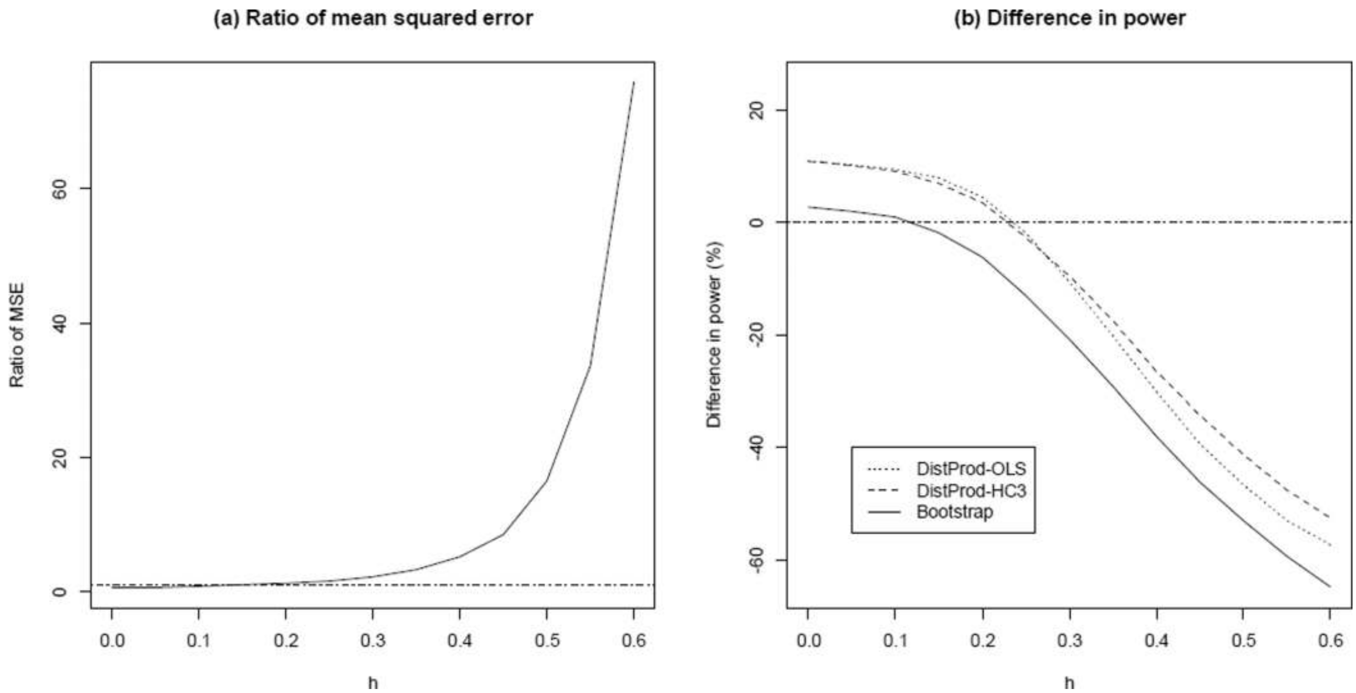

Figure 3.

Performance of standard mediation analysis vs. proposed robust mediation analysis when data are generated from the Tukey *g*-and-*h* family of distributions with different values of *h*. A larger value of *h* indicates heavier tails. Panel (a) shows the ratio of the mean squared error (MSE). For reference, a ratio of 1 is indicated by the dashed horizontal line. Panel (b) shows the difference in power. DistProd-OLS and DistProd-HC3 denote the distributions of the product methods based on the ordinary least squares (OLS) and heteroscedasticity consistent covariance (HC3) estimates of variance, respectively. A difference of 0 is indicated by the dashed horizontal line.

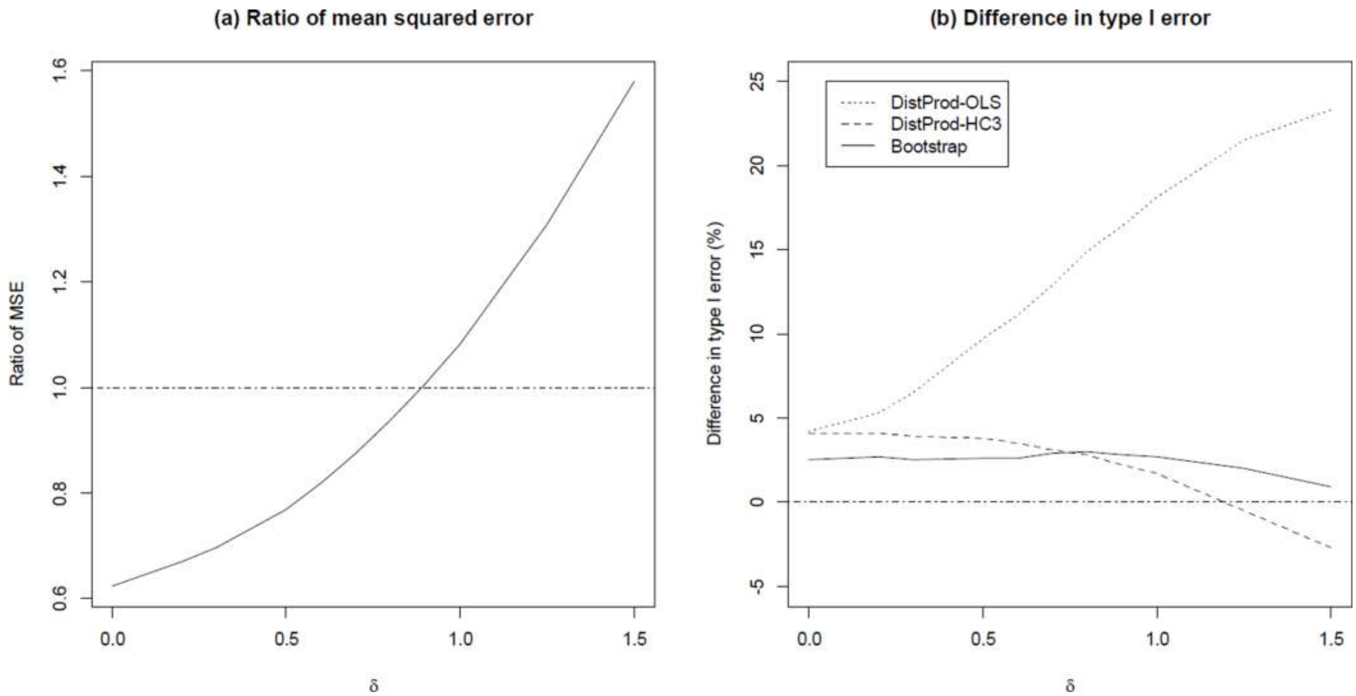

Figure 4.
Performance of standard mediation analysis vs. proposed robust mediation analysis under different values of δ where a larger value of δ indicates a higher degree of heteroscedasticity. Panel (a) shows the ratio of the mean squared error (MSE). For reference, a ratio of 1 is indicated by the dashed horizontal line. Panel (b) shows the difference in type I error rate. DistProd-OLS and DistProd-HC3 denote the distributions of the product methods based on the ordinary least squares (OLS) and heteroscedasticity consistent covariance (HC3) estimates of variance, respectively. A difference of 0 is indicated by the dashed horizontal line.

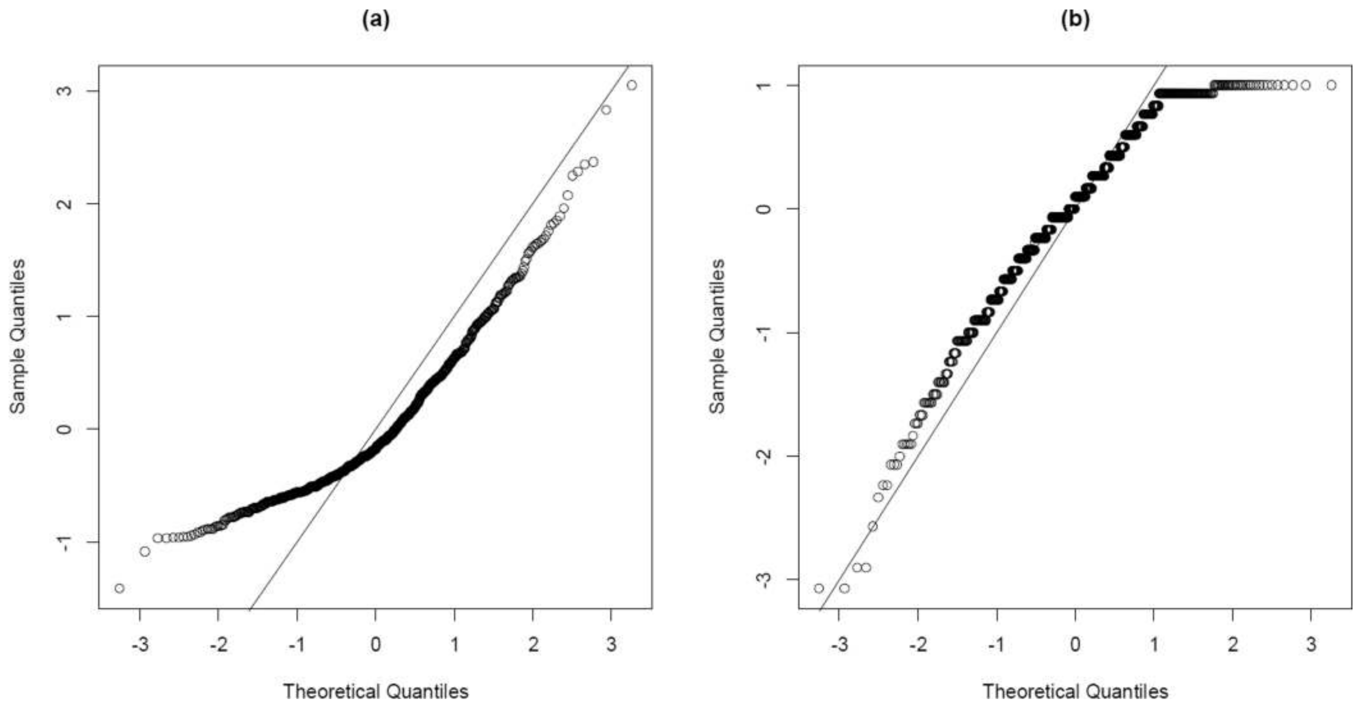

Figure 5.
Panel (a) is the normal quantile-quantile plot for the residuals of *y* regressing on *x* and *m*; panel (b) is the normal quantile-quantile plot for the residuals of *m* regressing on *x*.

Table 1

The relative root mean squared error (×1000) for estimates of the mediation effect using the standard ordinary least squares (OLS) and proposed robust approaches under different error distributions.

Error distribution	N	Small effect		Medium effect		Large effect	
		OLS	Robust	OLS	Robust	OLS	Robust
Normal	50	65	87	19	24	12	15
N(0,1)	100	40	54	13	16	8	11
	200	23	32	9	11	6	7
	500	16	16	5	7	3	4
	1000	0	16	4	5	2	3
Heavy-tailed	50	215	77	66	23	43	15
$t_{df=2}$	100	151	46	40	15	26	10
	200	84	28	27	10	18	6
	500	43	16	15	6	10	4
	1000	32	0	11	4	8	3
Contaminated	50	182	67	49	21	32	14
$0.9N(0,1)$ +	100	108	40	33	14	21	9
$0.1N(0, 10^2)$	200	40	16	21	9	14	6
	500	65	23	13	6	9	4
	1000	28	0	9	4	6	3
Heteroscedastic	50	123	109	31	28	20	18
$N(0, x_i^2)$	100	74	67	21	19	13	12
	200	43	40	13	13	9	8
	500	28	23	9	8	6	5
	1000	16	16	6	6	4	4

Table 2

Type I error rate of testing the mediation effect using the standard ordinary least squares (OLS) and proposed robust approaches under different error distributions.

Error distribution	N	Distribution of product			Bootstrap	
		OLS	HC3	Robust	OLS	Robust
Normal	50	2.8	2.7	0	6.6	0.6
N(0,1)	100	4.2	4.1	0.2	8	1.7
	200	5.4	5.3	1.2	7.4	3.3
	500	4.8	4.8	2.3	5.7	4.1
	1000	5.4	5.3	3.3	5.8	4.5
Heavy-tailed	50	1.1	0.5	0	1.4	0.5
$t_{df=2}$	100	1.3	1	0.1	2.2	1.5
	200	2.4	1.7	1.3	3.5	3.1
	500	3.3	2.5	3.5	4.8	4.1
	1000	4.2	3.1	5.1	5.8	4.8
Contaminated	50	0.8	0.4	0	1.4	0.7
0.9*N*(0,1) +	100	1.4	0.4	0.1	1.7	1.9
0.1*N*(0, 10^2)	200	1.6	0.6	0.4	2.4	3.5
	500	3.4	1.7	1.9	4.4	4
	1000	5	3.4	3	6.3	4.5
Heteroscedastic	50	15.9	2	0.8	2.8	0.7
$N(0, x_i^2)$	100	20.6	3.4	2.3	4.1	2
	200	21.8	5.4	3.7	6.3	3.6
	500	25.4	5.3	5.1	5.9	4
	1000	26.5	5.2	5	5.6	4.6

Note. HC3, heteroscedasticity consistent covariance

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 3

Power of the standard ordinary least squares (OLS) and proposed robust approaches under different error distributions.

Sample size	Small effect					Medium effect				
	Distribution of product			Bootstrap		Distribution of product			Bootstrap	
	OLS	HC3	Robust	OLS	Robust	OLS	HC3	Robust	OLS	Robust
Normal errors: $N(0,1)$										
50	1.7	1.5	0.0	4.6	0.9	46.8	42.3	0.9	57.5	24.0
100	5.0	4.8	0.1	10.4	3.1	88.1	86.4	20.9	91.7	67.3
200	19.2	18.7	1.5	30.7	11.6	99.9	99.9	89.0	99.9	97.9
500	72.7	72.3	25.6	81.2	51.3	100.0	100.0	100.0	100.0	100.0
1000	98.0	97.9	79.2	98.5	88.8	100.0	100.0	100.0	100.0	100.0
Heavy-tailed errors: $t_{df=2}$										
50	1.3	0.8	0.0	1.6	0.7	11.8	13.5	1.0	17.0	19.8
100	2.1	2.1	0.2	3.8	3.6	26.1	30.7	17.3	34.0	55.4
200	4.8	6.0	2.8	8.0	14.7	51.6	54.9	74.5	56.9	92.9
500	14.6	18.0	32.0	20.3	52.7	80.6	82.8	99.9	82.9	100.0
1000	30.2	34.5	73.4	36.3	84.0	93.1	93.3	100.0	92.9	100.0
Contaminated errors: $0.9N(0,1) + 0.1N(0, 10^2)$										
50	1.0	0.9	0.0	2.7	1.6	8.2	11.2	3.4	15.7	24.7
100	1.4	2.2	0.6	4.6	6.5	17.0	22.0	28.0	24.6	61.3
200	12.0	13.0	41.4	13.9	55.9	36.9	39.3	84.2	40.3	95.6
500	3.7	5.9	8.0	7.8	21.1	72.0	71.9	100.0	71.7	100.0
1000	25.4	25.3	80.3	26.0	86.2	94.5	93.7	100.0	93.4	100.0
Heteroscedastic errors: $N(0, x_i^2)$										
50	10.2	0.8	0.3	1.2	0.3	45.2	12.6	5.9	18.3	11.6
100	15.3	1.8	0.7	2.5	1.1	74.5	35.8	28.6	43.5	40.7
200	24.1	4.6	2.6	5.4	3.9	97.1	84.1	79.2	86.9	88.1
500	55.9	16.9	17.2	18.4	21.4	100.0	99.5	99.9	99.7	99.9
1000	84.5	47.3	54.1	48.9	59.5	100.0	100.0	100.0	100.0	100.0

Note. HC3, heteroscedasticity consistent covariance; the pattern of the results under the large effect size is similar to that under the medium effect size, thus not shown.

Table 4

Estimates of the mediation effect and the associated standard error (SE) and 95% confidence intervals using the standard and robust mediation methods for the job training data set.

Mediation Method	$\hat{\alpha}\hat{\beta}$	95% Confidence Interval: Distribution of product: OLS	95% Confidence Interval: Distribution of product: HC3	95% Confidence Interval: Bootstrap
Standard	−0.015	(−0.039, 0.008)	(−0.039, 0.007)	(−0.039, 0.006)
Robust	−0.045	(−0.089, −0.007)		(−0.058, 0.000)

Note. Distribution of the product methods based on the ordinary least squares (OLS) and heteroscedasticity consistent covariance (HC3) estimates of variance apply to only the standard mediation method.

Table 5

The relative root mean squared error (×1000) for estimates of the average mediation effect using the maximum likelihood estimate (MLE) method and proposed robust approach for multilevel models with different error distributions when the effect size is small (i.e., α = β = 0.3).

		Number of level-two units (N)							
		25		50		100		200	
Error distribution	n_j	MLE	Robust	MLE	Robust	MLE	Robust	MLE	Robust
Normal	4	20	22	13	16	9	11	7	8
$N(0,1)$	8	16	16	11	12	8	9	5	6
	16	14	14	10	10	7	7	5	5
Heavy-tailed	4	131	25	84	18	55	13	37	9
$t_{df=2}$	8	92	19	51	14	35	9	26	7
	16	54	16	35	11	24	8	18	5
Contaminated	4	51	24	32	16	22	12	15	9
$0.9N(0,1)$ +	8	34	18	23	13	16	9	11	6
$0.1N(0, 10^2)$	16	24	16	17	11	12	8	8	5
Heteroscedastic	4	27	27	18	19	12	13	9	10
$N(0, x_i^2)$	8	20	19	14	14	10	10	7	7
	16	17	16	12	11	9	8	6	6

NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 6

Power of the maximum likelihood estimate (MLE) method and proposed robust approach for multilevel models with different error distributions when the effect size is small (i.e., α = β = 0.3).

		Number of level-two units (N)							
		25		50		100		200	
Error distribution	n_j	MLE	Robust	MLE	Robust	MLE	Robust	MLE	Robust
Normal	4	19.7	5.2	66.3	38.0	99.0	91.8	100.0	100.0
$N(0,1)$	8	44.2	27.4	91.6	82.0	100.0	99.8	100.0	100.0
	16	63.4	65.9	98.6	97.8	100.0	100.0	100.0	100.0
Heavy-tailed	4	0.2	8.0	0.2	37.7	1.8	80.0	7.6	99.1
$t_{df=2}$	8	0.5	26.0	1.9	72.4	6.4	98.6	25.2	100.0
	16	1.6	56.5	6.7	95.3	20.4	100.0	48.7	100.0
Contaminated	4	1.1	6.9	3.4	36.3	18.8	85.9	56.5	99.5
$0.9N(0,1)$ +	8	3.6	28.1	16.4	79.7	52.3	99.9	88.6	100.0
$0.1N(0, 10^2)$	16	13.1	63.3	45.6	97.1	86.4	100.0	99.5	100.0
Heteroscedastic	4	7.4	5.0	29.1	27.5	79.3	80.2	99.7	99.1
$N(0, x_i^2)$	8	16.8	19.9	60.9	69.8	97.8	97.9	100.0	100.0
	16	33.6	51.0	81.8	92.9	100.0	100.0	100.0	100.0