Manuscript (2020), **xx**, x, *pp.* 1–15

Nonequivalence of two least-absolute-deviation estimators for mediation effect

BY WENWU WANG
School of Statistics, Qufu Normal University, Jingxuan West Road, Qufu 273165, China
wengewsh@sina.com

PING YU
HKU Business School, University of Hong Kong, Pokfulam Road, Hong Kong
pingyu@hku.hk

SUMMARY

This paper provides two groups of conditions of model consistency in least-absolute-deviation mediation models. Under model consistency, we shows that the difference estimator and the product estimator are not only numerically nonequivalent but asymptotically nonequivalent, which is dramatically different from the situation in the least squares mediation analysis where these two estimators are numerically equivalent. In all three possible scenarios of model parameters, both the asymptotic theories and simulation studies show that the product estimator is more efficient than the difference estimator.

Some key words: LAD mediation; Model consistency; Difference estimator; Product estimator; Second-order asymptotic.

1. INTRODUCTION

In many disciplines, the effect of the predictor on the outcome variable is often affected by a third variable termed as a mediator. Mediation analysis aims to identify the mediation effect between the predictor and the outcome via the change in the mediator (MacKinnon, 2008; VanderWeele, 2015; Hayes, 2018). Since the seminal paper of Baron & Kenny (1986), the empirical applications of mediation analysis have dramatically expanded in sociology, psychology, epidemiology, and medicine. For example, Lindquist (2012) determined whether activation in certain brain regions mediated the effect of applied temperature on self-reported pain based on the data from a functional magnetic resonance imaging study of thermal pain. VanderWeele et al. (2013) examined whether the 4Rs intervention has an effect on students' depressive symptoms by changing the quality of other classes and changing the quality of a student's own class.

The basic mediation model consists of three regression equations with the causality assumptions. One usually adopts the least squares (LS) or the maximum likelihood under normality to obtain two estimates of the mediation effect: the difference estimate and the product estimate (MacKinnon et al., 2002). These two estimates are numerically equivalent under mild conditions (MacKinnon et al., 1995; Wang et al., 2020).

Unfortunately, the real data sets are seldom normally distributed. Micceri (1989) examined 440 data sets from the psychological and educational literature and found that none of them was normally distributed. Instead, their distributions were either heavy-tailed and/or skewed. Field & Wilcox (2017) further showed that some assumptions about commonly used statistical methods are poorly understood and likely to be violated in psychological data, so inappropriate statistical methods are often applied. Mediation analysis for non-normal variables is an active research area nowadays (Preacher, 2015).

It is well known that the LS method may break down in the presence of outliers and heavy-tailed errors. Therefore, it is important to develop robust alternatives of the LS method. Zu & Yuan (2010) used the local influence method to identify outliers which are far away from the majority of observations and strongly affect the mediation analysis. Recently, Yuan & MacKinnon (2014) proposed the product estimate of the mediation effect by applying the least-absolute-deviation (LAD) regression to data with heavy-tailed errors. Shen et al. (2014) and Bind et al. (2017) suggested to employ quantile mediation models to explore the information in the error distribution.

The purpose of this paper is to provide an rigorous analysis on model consistency and show nonequivalence of the difference and product estimators in the LAD mediation model. We provide two groups of conditions of model consistency. Both groups of conditions allow heavy-tailed errors; the first group allows heteroskedasticity but the errors must be symmetrically distributed about zero, while the second group excludes heteroskedasticity but allows skewed errors. Under model consistency, we show that the two LAD estimators are not only numerically nonequivalent but asymptotically nonequivalent, which is dramatically different from the LS case where the two estimators are numerically equivalent. Specifically, for all three possible scenarios of model parameters, our asymptotic theories show that the product estimator is more efficient than the difference estimator, which is confirmed by our simulation studies.

2. LAD MEDIATION ANALYSIS

2.1. *LAD mediation model*

As mentioned in the Introduction, the mean (or the LS method) is sensitive to outliers and performs poorly when the error distribution is heavy-tailed (Huber & Ronchetti, 2009; Wilcox, 2017). Thus, it is not always an appropriate summary of the center of the data distribution; as an alternative, the median may perform better for non-normal distributions with heavy-tails. In this section, we apply the LAD regression to the basic mediation model, give two conditions for model consistency, and provide the asymptotic theory for the LAD estimates of the basic parameters of the model.

Given the observations $(X_i, M_i, Y_i), i = 1, \ldots, n$, the basic LAD mediation model consists of three regression equations:

$$Y_i = \beta_1 + cX_i + \epsilon_{1i} \tag{1}$$
$$M_i = \beta_2 + aX_i + \epsilon_{2i} \tag{2}$$
$$Y_i = \beta_3 + c'X_i + bM_i + \epsilon_{3i}, \tag{3}$$

where the errors satisfy $\text{Med}(\epsilon_{ki} \mid X_i) = 0$ for $k = 1, 2$, and $\text{Med}(\epsilon_{3i} \mid X_i, M_i) = \text{Med}(\epsilon_{3i} \mid X_i, \epsilon_{2i}) = 0$, which implies $\text{Med}(\epsilon_{3i} \mid X_i) = 0$. To compare with Equation (1), we plug Equation (2) into Equation (3) to have

$$Y = \beta_3 + c'X + b\left(\beta_2 + aX + \epsilon_2\right) + \epsilon_3 = \left(\beta_3 + b\beta_2\right) + \left(c' + ab\right)X + \epsilon_i. \tag{4}$$

where $\epsilon_i = b\epsilon_{2i} + \epsilon_{3i}$. The parameter c represents the total effect of X on Y, a measures the relation between X and M, c' represents the direct effect of X on Y adjusted for the effect of M, and b measures the relation between M and Y adjusted for the effect of X.

In the general setup of the LAD mediation model, we only assume that the conditional medians of the errors are zero, and do not impose any other distributional assumptions on them, i.e., independence (of X), homoscedasticity or normality; see Section 2.2 for further discussions on the error term distributions. As in the LS mediation model, the LAD mediation effect can be defined in two forms: the product of parameters ab and the difference in parameters $c - c'$. Next, we discuss the conditions of model consistency, i.e., when will the difference in parameters $c - c'$ be equal to the product of parameters ab in population?

2.2. *Model consistency*

In LAD mediation analysis, Yuan & MacKinnon (2014) discussed the model consistency for normally distributed errors in their Theorem 1. We restated their Theorem 1 using our own language in the following Lemma 1.

LEMMA 1. *If conditional on X_i, ϵ_{2i} and ϵ_{3i} are independent and zero-mean normally distributed, then* $\text{Med}(b\epsilon_{2i} + \epsilon_{3i} \mid X_i) = 0$ *such that $c - c' = ab$ holds.*

Note that we do not allow dependency between ϵ_{2i} and ϵ_{3i} conditional on X_i. This is because if their correlation corr$(\epsilon_{2i}, \epsilon_{3i} \mid X_i) \equiv \rho_{23}(X_i) \neq 0$, then $\text{Med}(\epsilon_{3i} \mid X_i, \epsilon_{2i}) = E(\epsilon_{3i} \mid X_i, \epsilon_{2i}) = \epsilon_{2i}\rho_{23}(X_i)\sigma_3(X_i)/\sigma_2(X_i) \neq 0$, violating the basic requirement of Equation (3) in LAD mediation analysis, where $\sigma_k(X_i)$ is the conditional standard deviation of ϵ_{ki} given X_i, $k = 2, 3$. As a result, the dependency between ϵ_{2i} and ϵ_{3i} can only be through X_i. A typical case satisfying the assumptions in Lemma 1 is $\epsilon_{ki} = \sigma_k(X_i)\varepsilon_{ki}$ with ε_{2i} and ε_{3i} being independent and following $N(0, 1)$.

Lemma 1 is a good starting point for the model consistency in LAD mediation analysis. However, the normality of errors is too strong as it is well known that the LS estimator is more efficient than the LAD estimator for normal errors. The following proposition tries to extend the model consistency to non-normal errors in two typical cases; the first case allows dependency of $(\epsilon_{2i}, \epsilon_{3i})$ on X_i while the second case does not.

PROPOSITION 1. *If either of the following two cases hold, (i) conditional on X_i, ϵ_{2i} and ϵ_{3i} are independent and symmetrically distributed about* 0*; (ii)* $(\epsilon_{2i}, \epsilon_{3i})$ *are independent of X_i, and* $\text{Med}(\epsilon_{2i}) = 0$ *and* $\text{Med}(\epsilon_{3i} \mid \epsilon_{2i}) = 0$*; then in LAD mediation analysis, we have the equality*

$$c - c' = ab.$$

The conditions in Case (i) imply $\text{Med}(\epsilon_{2i} \mid X_i) = 0$ and $\text{Med}(\epsilon_{3i} \mid X_i, \epsilon_{2i}) = \text{Med}(\epsilon_{3i} \mid X_i) = 0$, and the conditions in Case (ii) imply $\text{Med}(\epsilon_{2i} \mid X_i) = \text{Med}(\epsilon_{2i}) = 0$ and $\text{Med}(\epsilon_{3i} \mid \epsilon_{2i}, X_i) = \text{Med}(\epsilon_{3i} \mid \epsilon_{2i}) = 0$, which are the requirements of the LAD mediation model, so are not restrictions. The real restrictions are independence between ϵ_{2i} and ϵ_{3i} and the symmetricity of their distributions given X_i in Case (i) and the independence between $(\epsilon_{2i}, \epsilon_{3i})$ and X_i in Case (ii). Case (i) in Proposition 1 is an extension of Lemma 1. It implies that the normality is not the key for $\text{Med}(b\epsilon_{2i} + \epsilon_{3i} \mid X_i) = 0$, but the symmetricity (implied by the normality) is. A typical scenario in Case (i) is $\epsilon_{ki} = \sigma_k(X_i)\varepsilon_{ki}$, $k = 2, 3$, with ε_{2i} and ε_{3i} being independent and symmetrically distributed about 0. Compared with Case (i), Case (ii) requires the independence between $(\epsilon_{2i}, \epsilon_{3i})$ and X_i but relaxes the independence between ϵ_{2i} and ϵ_{3i} and the symmetricity of their distributions, which implies that when the means of ϵ_{2i} and ϵ_{3i} exist, $E(\epsilon_{2i})$ and $E(\epsilon_{3i})$ need not be zero. From the proof of Proposition 1 in Case (ii), $\text{Med}(\epsilon_i)$ need not be zero such that β_1 need not be equal to $\beta_3 + b\beta_2$, but $c - c' = ab$ still holds. When $\mathrm{Med}(\epsilon_i) \neq 0$, we denote $d = \mathrm{Med}(\epsilon_i)$, and then $\beta_1 = d + \beta_3 + b\beta_2$ and $\epsilon_{1i} = \epsilon_i - d$. The conditions in Case (i) of Proposition 1 are hard to relax, as shown in the following two examples.

Example 1. This example will show that in Case (i), even if ϵ_{2i} and ϵ_{3i} are symmetrically distributed about 0 given X_i, ϵ_i conditional on X_i need not be symmetrically distributed about 0 when ϵ_{2i} and ϵ_{3i} are correlated. As a result, $\mathrm{Med}(\epsilon_i \mid X_i)$ will depend on X_i and $c - c' = ab$ cannot hold. Assume $\epsilon_{ki} = \sigma(X_i)\,\varepsilon_{ki}$, $k = 2, 3$, where ε_{2i} and ε_{3i} are symmetrically distributed about 0 but correlated with each other. Specifically, if $(\varepsilon_{2i}, \varepsilon_{3i})$ take three values $(-1, 1), (0, -1), (1, 1)$ with probabilities $1/4, 1/2, 1/4$ respectively, then both ε_{2i} and ε_{3i} are symmetrically distributed about 0, but $\varepsilon_{2i} + \varepsilon_{3i}$, which is ε_i when $b = 1$, is not. In consequence, $\mathrm{Med}(\epsilon_i \mid X_i) = \mathrm{Med}(\varepsilon_i)\sigma(X_i) = d'\sigma(X_i)$ depending on X_i, and $c - c' = ab$ cannot hold, where $d' = \mathrm{Med}(\varepsilon_i)$. Of course, independence between ε_{2i} and ε_{3i} are only sufficient but not necessary for the symmetricity of $\varepsilon_{2i} + \varepsilon_{3i}$. For example, if $(\varepsilon_{2i}, \varepsilon_{3i})$ are uniformly distributed on the unit disc (and hence have symmetric marginal densities but are not independent), then $\varepsilon_{2i} + \varepsilon_{3i}$ also has a symmetric density.

Example 2. This example will show that in Case (i), even if ϵ_{2i} and ϵ_{3i} are independent given X_i, ϵ_i conditional on X_i need not be symmetrically distributed about 0 if either ϵ_{2i} or ϵ_{3i} is not symmetrically distributed about 0. As a result, $\mathrm{Med}(\epsilon_i \mid X_i)$ will depend on X_i and $c - c' = ab$ cannot hold. Suppose X takes three values, $0, 1$ and 2, $\epsilon_2 \sim N(0, 1)$, $\epsilon_3 \sim \lambda_X\{Exp(1) - \ln 2\}$ with the scale parameter $\lambda_X = X + 1$, and $b = 1$, where $N(0, 1)$ and the standard exponential distribution $Exp(1)$ are independent. By numerical simulation, we obtain three different medians $\gamma_0 = 0.88$, $\gamma_1 = 1.58$ and $\gamma_2 = 2.24$ which satisfy

$$pr(\epsilon_1 \leq \gamma_x) = pr(\epsilon_3 \leq \gamma_x - b\epsilon_2) = 0.5,$$

where $x = 0, 1, 2$. Obviously, the three medians are not the same and $c - c' = ab$ fails.

In Example 2, γ_0, γ_1 and γ_2 do not fall on a straight line. In Example 1, if $\sigma(X_i)$ is not affine in X_i (i.e., $\sigma(X_i)$ cannot be expressed as $a + bX_i$ for some real numbers a and b), $\mathrm{Med}(\epsilon_i \mid X_i)$ is not affine in X_i. As a result, we cannot even find β_1 and c in Equation (1) such that $\mathrm{Med}(\epsilon_{1i} \mid X_i) = 0$. Of course, we can treat $\beta_1 + cX_i$ as an approximation of the true (nonlinear) conditional median $\mathrm{Med}(Y_i|X_i)$ as in Angrist et al. (2006), but this seems out of the scope of the usual mediation analysis. In contrast, in the LS mediation analysis, as long as $E(\epsilon_{2i} \mid X_i) = E(\epsilon_{3i} \mid X_i) = 0$, $E(\epsilon_i \mid X_i) = E(b\epsilon_{2i} + \epsilon_{3i} \mid X_i) = 0$. In summary, although the LS estimate of the mediation effect is less robust to the heavy-tailedness of error distributions, its model consistency allows a less restrictive relationship between $(\epsilon_{2i}, \epsilon_{3i})$ and X_i and/or the joint distribution of $(\epsilon_{2i}, \epsilon_{3i})$. So there is a trade-off between the choices of the LS and LAD mediation models.

2.3. *LAD estimates of mediation effect*

The LS method is to minimize the squared errors between the dependent variable and the regression function, while the LAD method is to minimize the absolute values of errors. Compared with the sum of the squared errors, the sum of the absolute values of errors is not sensitive to outliers. Thus, the LAD is a useful alternative to the LS when facing outliers or heavy-tailed errors. The LAD estimates of regression parameters in Equations (1)-(3) are obtained by

$$(\widehat{\beta}_1, \widehat{c}) = \arg\min_{\beta_1, c} \sum_{i=1}^{n} |Y_i - \beta_1 - cX_i| ,$$

$$(\widehat{\beta}_2, \widehat{a}) = \arg\min_{\beta_2, a} \sum_{i=1}^{n} |M_i - \beta_2 - aX_i| ,$$

$$(\widehat{\beta}_3, \widehat{c'}, \widehat{b}) = \arg\min_{\beta_3, c', b} \sum_{i=1}^{n} \left|Y_i - \beta_3 - c'X_i - bM_i\right| .$$

For large samples, the LAD estimate is approximately normally distributed (Pollard, 1991; Koenker, 2005). To rigorously state the asymptotic distributions of $\widehat{c}$, $\widehat{c'}$, $\widehat{a}$ and $\widehat{b}$, which are the building blocks of our parameters of interest $\widehat{c} - \widehat{c'}$ and $\widehat{a}\widehat{b}$, we impose the following conditions. First, for a generic random variable ϵ, we define f_ϵ as the probability density function (pdf) of ϵ.

Condition 1. (i) x, ϵ_2 and ϵ_3 are independent of each other. (ii) $\mathrm{Med}\,(\epsilon_2) = 0$ and $\mathrm{Med}(\epsilon_3) = 0$. (iii) $f_{\epsilon_k}(0) > 0$ for $k = 1, 2, 3$. (iv) $E(|x|^2) < \infty$ and $E(|\epsilon_2|^2) < \infty$. (v) f_{ϵ_1}, f_{ϵ_2} and f_{ϵ_3} are continuous at 0.

Condition 1 collects the conditions for the first-order expansions of $\widehat{c}$, $\widehat{c'}$, $\widehat{a}$ and $\widehat{b}$. Compared with Case (ii) of Proposition 1, Condition 1 restricts ϵ_2 and ϵ_3 to be independent; anyway, such an restriction allows heavy-tailedness and skewness of ϵ_2 and ϵ_3 as encountered in practice. This restriction would greatly simplify our analysis; otherwise, tedious formulae would blur the main conclusions of our analysis. Assuming errors of a causal diagram to be independent is not new; see, e.g., Pearl (1995, 2009). Under this restriction, $F_{\epsilon_1}(x) = \int F_{\epsilon_3}(x + d - b\epsilon_2) f_{\epsilon_2}(\epsilon_2)\mathrm{d}\epsilon_2$ and $f_{\epsilon_1}(x) = \int f_{\epsilon_3}(x + d - b\epsilon_2) f_{\epsilon_2}(\epsilon_2)\mathrm{d}\epsilon_2$, where for a generic random variable ϵ, $F_\epsilon(\cdot)$ is the cumulative distribution function (cdf) of ϵ. Given the independence between ϵ_2 and ϵ_3, the nonzeroness of μ_{ϵ_2}, μ_{ϵ_3} and d is only from the asymmetricity of the distributions of ϵ_2 and ϵ_3, where for a generic random variable X, $\mu_X = E(X)$. When ϵ_2 and ϵ_3 are symmetrically distributed, $d = 0$, $F_{\epsilon_3}(\epsilon_3) = 1 - F_{\epsilon_3}(-\epsilon_3)$ and $f_{\epsilon_3}(\epsilon_3) = f_{\epsilon_3}(-\epsilon_3)$, so $F_{\epsilon_1}(x) = \int \{1 - F_{\epsilon_3}(b\epsilon_2 - x)\} f_{\epsilon_2}(\epsilon_2)\mathrm{d}\epsilon_2$ and $f_{\epsilon_1}(x) = \int f_{\epsilon_3}(b\epsilon_2 - x) f_{\epsilon_2}(\epsilon_2)\mathrm{d}\epsilon_2$. Other conditions in Condition 1 are standard regularity assumptions.

From Pollard (1991) and Knight (1998), we have the following asymptotic distributions for $\widehat{c}$, $\widehat{c'}$, $\widehat{a}$ and $\widehat{b}$.

LEMMA 2. *Under Condition* 1*,*

$$n^{1/2}(\widehat{c} - c) \Longrightarrow N\left(0, \frac{H^{2,2}}{4f_{\epsilon_1}^2(0)}\right), \quad n^{1/2}(\widehat{c'} - c') \Longrightarrow N\left(0, \frac{J^{2,2}}{4f_{\epsilon_3}^2(0)}\right),$$

$$n^{1/2}(\widehat{a} - a) \Longrightarrow N\left(0, \frac{H^{2,2}}{4f_{\epsilon_2}^2(0)}\right), \quad n^{1/2}(\widehat{b} - b) \Longrightarrow N\left(0, \frac{J^{3,3}}{4f_{\epsilon_3}^2(0)}\right),$$

where the symbol '$\Longrightarrow$' signifies weak convergence of the associated probability measures, for a matrix A, $A^{i,j}$ is the (i, j) element of A^{-1}, $H = \mathrm{E}(\mathbf{x}_1\mathbf{x}_1^{\mathrm{T}})$ with $\mathbf{x}_1 = (1, X)^{\mathrm{T}}$, $J = \mathrm{E}(\mathbf{x}_2\mathbf{x}_2^{\mathrm{T}})$, with $\mathbf{x}_2 = (1, X, M)^{\mathrm{T}}$, and the superscript 'T' indicates the transpose of a matrix.

For a generic random variable X, define σ_X^2 =var(X), and set σ_k^2 =var(ϵ_k), $k = 1, 2, 3$. Under Condition 1, if σ_k^2 exists, then we need only replace $(4f_{\epsilon_k}^2(0))^{-1}$ in Lemma 2 by σ_k^2 to get the asymptotic distributions of the corresponding LS estimates. Note that the LAD estimates need not dominate the LS estimates in efficiency. From Lemma 2, the former is more efficient than the latter if and only if the ratio $\sigma_k^2/(4f_{\epsilon_k}^2(0))^{-1} > 1$. For example, this ratio is $0.64, 5.75, +\infty$ for the standard normal distribution $N(0,1)$, the contaminated normal distribution $0.9N(0,1) + 0.1N(0,10^2)$, and the heavy-tailed distribution t_2 which is the t-distribution with two degrees of freedom, respectively.

Based on the LAD estimators of $\widehat{c}$, $\widehat{c'}$, $\widehat{a}$ and $\widehat{b}$, we can construct the difference estimator $\widehat{c} - \widehat{c'}$ and the the product estimator $\widehat{ab}$ of the mediation effect, where the latter estimator is suggested by Yuan & MacKinnon (2014) in the LAD mediation model. Unlike the numerical equivalence of the two parallel LS estimates (MacKinnon et al., 1995; Wang et al., 2020), the two LAD estimates are not numerically equivalent, that is, $\widehat{ab} \neq \widehat{c} - \widehat{c'}$ in general (see our simulation studies in Section 4). Actually, they are not even asymptotically equivalent, as shown in the coming section.

3. ASYMPTOTIC THEORY FOR LAD ESTIMATES OF MEDIATION EFFECT

3.1. *Conditions for second-order asymptotics*

We now study the asymptotic properties of $\widehat{c} - \widehat{c'}$, $\widehat{ab}$ and $\widehat{c} - \widehat{c'} - \widehat{ab}$, where $\widehat{c}, \widehat{c'}, \widehat{a}$ and $\widehat{b}$ are the LAD estimators defined in Section 2.3. It turns out that these asymptotic properties critically depend on the zeroness of a and b. Whether $b = 0$ or not determines $\widehat{c} - \widehat{c'} - \widehat{ab}$ is $n^{1/2}$-consistent or $n^{3/4}$-consistent. Given $b = 0$, whether $a = 0$ or not determines the convergence rates of $\widehat{c} - \widehat{c'}$ and $\widehat{ab}$ to be $n^{1/2}$ or faster than $n^{1/2}$. For $n^{1/2}$-consistency, we require only the first-order expansions of $\widehat{c} - \widehat{c'}$ and $\widehat{ab}$, while for $n^{3/4}$-consistency, we require their second-order expansions. The second-order expansions need stronger conditions than the first-order expansions as detailed below.

Condition 2. (i)-(iii) are the same as in Condition 1. (iv) $E(|x|^3) < \infty$ and $E(|\epsilon_2|^3) < \infty$. (v) f_{ϵ_2} is continuous at 0, and f_{ϵ_3} is differentiable at 0.

Compared with Condition 1, Condition 2 imposes stronger conditions on the moments of x and ϵ_2 and the smoothness of f_{ϵ_3} at 0. Because $\epsilon_1 = \epsilon_3$ when $b = 0$, $f_{\epsilon_1}(0) > 0$ and f_{ϵ_1} is continuous at 0 in Condition 1 can be omitted. Now, we are ready to discuss the asymptotic properties of $\widehat{c} - \widehat{c'}$, $\widehat{ab}$ and $\widehat{c} - \widehat{c'} - \widehat{ab}$ when $b \neq 0$, $b = 0$ but $a \neq 0$, and $a = b = 0$ which are labeled as Case one, Case two and Case three, respectively.

3.2. *Case one: $b \neq 0$*

We first consider the case with $b \neq 0$ in the following Theorem 1.

THEOREM 1. *When $b \neq 0$, if Condition 1 holds, then*

$$n^{1/2}\{\widehat{c} - \widehat{c'} - (c - c')\} \Longrightarrow N(0, \sigma_D^2), \quad n^{1/2}(\widehat{ab} - ab) \Longrightarrow N(0, \sigma_P^2),$$
$$n^{1/2}(\widehat{c} - \widehat{c'} - \widehat{ab}) \Longrightarrow N(0, \sigma^2),$$

where

$$\sigma_D^2 = \frac{1}{4f_{\epsilon_1}(0)^2\sigma_X^2} + \frac{1/\sigma_X^2 + a^2/\sigma_2^2}{4f_{\epsilon_3}(0)^2} - 2\frac{\Lambda(\epsilon_1,\epsilon_3)}{\sigma_X^2}, \quad \sigma_P^2 = \frac{b^2}{4f_{\epsilon_2}(0)^2\sigma_X^2} + \frac{a^2}{4f_{\epsilon_3}(0)^2\sigma_2^2},$$
$$\sigma^2 = \frac{1}{\sigma_X^2}\left[\frac{1}{4f_{\epsilon_1}(0)^2} + \frac{1}{4f_{\epsilon_3}(0)^2} + \frac{b^2}{4f_{\epsilon_2}(0)^2} - 2\Lambda(\epsilon_1,\epsilon_3) - 2b\Lambda(\epsilon_1,\epsilon_2)\right],$$

with

$$\Lambda(\epsilon_1, \epsilon_k) = \frac{E\{1(\epsilon_1 \le 0)1(\epsilon_k \le 0)\} - 1/4}{f_{\epsilon_1}(0) f_{\epsilon_k}(0)}, k = 2, 3,$$

$$E\{1(\epsilon_1 \le 0)1(\epsilon_2 \le 0)\} = \int_{-\infty}^{0} F_{\epsilon_3}(-b\epsilon_2 + d)\mathrm{d}F_{\epsilon_2}(\epsilon_2),$$

$$E\{1(\epsilon_1 \le 0)1(\epsilon_3 \le 0)\} = \begin{cases} \frac{1}{2}F_{\epsilon_2}(\frac{d}{b}) + \int_{d/b}^{\infty} F_{\epsilon_3}(-b\epsilon_2 + d)\mathrm{d}F_{\epsilon_2}(\epsilon_2), & \textit{if } b > 0, \\ \frac{1}{2}(1 - F_{\epsilon_2}(\frac{d}{b})) + \int_{-\infty}^{d/b} F_{\epsilon_3}(-b\epsilon_2 + d)\mathrm{d}F_{\epsilon_2}(\epsilon_2), & \textit{if } b < 0. \end{cases}$$

We give a few comments on Theorem 1 here. First, the asymptotic variance of $n^{1/2}(\widehat{c} - \widehat{c'} - \widehat{ab})$ is the not the sum of those of $n^{1/2}\{\widehat{c} - \widehat{c'} - (c - c')\}$ and $n^{1/2}(\widehat{ab} - ab)$ because they are asymptotically correlated. Their asymptotic covariance is $b\Lambda(\epsilon_1, \epsilon_2)$ which is included in the last term of σ^2. Second, the asymptotic distributions of $\widehat{c} - \widehat{c'}$, $\widehat{ab}$ and $\widehat{c} - \widehat{c'} - \widehat{ab}$ do not depend on β_2, β_3 and c', and that of $\widehat{c} - \widehat{c'} - \widehat{ab}$ does not even depend on a. Third, when $d = 0$,

$$E\{1(\epsilon_1 \le 0)1(\epsilon_3 \le 0)\} - \frac{1}{4} = \int_{0}^{\infty} F_{\epsilon_3}(-|b|\,\epsilon_2) dF_{\epsilon_2}(\epsilon_2),$$

so the formulae of σ_D^2 and σ^2 can be simplified. Fourth, when $a = 0$, the formulae of σ_D^2 and σ_P^2 can be simplified. However, σ^2 depends only on b but not on a. When $b = 0$, $\epsilon_1 = \epsilon_3$, so

$$\sigma^2 = \frac{1}{\sigma_X^2}\left[\frac{1}{4f_{\epsilon_3}(0)^2} + \frac{1}{4f_{\epsilon_3}(0)^2} - 2\frac{E\{1(\epsilon_3 \le 0)\} - 1/4}{f_{\epsilon_3}(0)^2}\right] = 0,$$

and a further refinement on the asymptotic distribution of $\widehat{c} - \widehat{c'} - \widehat{ab}$ is required; see Section 3.3. Fifth, it is interesting to discuss why the product estimator and the difference estimator are asymptotically (even numerically) equivalent in the LS mediation model while they are not in the LAD mediation model. From the proof of Theorem 1, the first-order asymptotic representations (FOARs) of the two estimators in the LAD mediation model are

$$\begin{aligned} n^{1/2}\{\widehat{c} - \widehat{c'} - (c - c')\} &= n^{-1/2}\sum_{i=1}^{n}\left\{\frac{\widetilde{X}_i}{\sigma_X^2}s(\epsilon_{1i}) - \left(\frac{\widetilde{X}_i}{\sigma_X^2} - \frac{a\widetilde{\epsilon}_{2i}}{\sigma_2^2}\right)s(\epsilon_{3i})\right\}, \\ n^{1/2}(\widehat{ab} - ab) &= n^{-1/2}\sum_{i=1}^{n}\left\{\frac{b\widetilde{X}_i}{\sigma_X^2}s(\epsilon_{2i}) + \frac{a\widetilde{\epsilon}_{2i}}{\sigma_2^2}s(\epsilon_{3i})\right\}, \end{aligned} \tag{5}$$

where $\widetilde{X}_i = X_i - \mu_X$, $\widetilde{\epsilon}_{2i} = \epsilon_{2i} - \mu_{\epsilon_2}$, and $s(\epsilon_{ki}) = \{1/2 - 1(\epsilon_{ki} \le 0)\}/f_{\epsilon_k}(0),\ k = 1, 2, 3,$ the two terms in $n^{1/2}\{\widehat{c} - \widehat{c'} - (c - c')\}$ are correlated even if ϵ_2 and ϵ_3 are independent, while the two terms in $n^{1/2}(\widehat{ab} - ab)$ are uncorrelated. As a result, the FOAR of $\widehat{c} - \widehat{c'} - \widehat{ab}$ is

$$n^{1/2}(\widehat{c} - \widehat{c'} - \widehat{ab}) = n^{-1/2}\frac{1}{\sigma_X^2}\sum_{i=1}^{n}\widetilde{X}_i\{s(\epsilon_{1i}) - s(\epsilon_{3i}) - bs(\epsilon_{2i})\},$$

which is not zero. In LS mediation analysis, we need only replace $s(\epsilon_{ki})$ by ϵ_{ki}, $k = 1, 2, 3$, to have the FOAR of $\widehat{c} - \widehat{c'} - \widehat{ab}$ as

$$n^{1/2}(\widehat{c} - \widehat{c'} - \widehat{ab}) = n^{-1/2}\frac{1}{\sigma_X^2}\sum_{i=1}^{n}\widetilde{X}_i\left(\epsilon_{1i} - \epsilon_{3i} - b\epsilon_{2i}\right) = 0.$$

where the last equality is from $\epsilon_{1i} = b\epsilon_{2i} + \epsilon_{3i}$.

Our simulation studies in Section 4 show that $\sigma_D^2 > \sigma_P^2$ for a few error distributions, i.e., the product estimator is more efficient. The following Corollary 1 rigorously states this fact when ϵ_2 and ϵ_3 follow the standard normal.

COROLLARY 1. *Under the assumptions of Theorem* 1*, if ϵ_2 and ϵ_3 both follow $N(0,1)$, then*

$$\sigma_D^2 - \sigma_P^2 = \{\pi - (b^2+1)^{1/2}(\pi - 2\arctan|b|)\}/\sigma_X^2,$$

which is positive when $b \neq 0$ and converges to $(\pi - 2)/\sigma_X^2$ when $|b| \to \infty$.

In Corollary 1, only the relative variance between ϵ_2 and ϵ_3 is relevant, and resetting $\epsilon_2 \sim N(0,\kappa^2)$ is equivalent to set b as $b\kappa$, so now $\sigma_D^2 - \sigma_P^2 = \left\{\pi - (b^2\kappa^2+1)^{1/2}(\pi - 2\arctan|b|\kappa)\right\}/\sigma_X^2$.

3.3. *Case two: $b = 0$ but $a \neq 0$*

We next consider the case with $b = 0$ but $a \neq 0$ in the following Theorem 2.

THEOREM 2. *When $b = 0$ but $a \neq 0$, if Condition* 1 *holds, then*

$$n^{1/2}(\widehat{c} - \widehat{c'}) \Longrightarrow N(0,\sigma_C^2), \quad n^{1/2}\widehat{ab} \Longrightarrow N(0,\sigma_C^2),$$

with $\sigma_C^2 = a^2/\{4f_{\epsilon_3}(0)^2\sigma_2^2\}$, and if further assume Condition 2 *holds, then,*

$$n^{3/4}(\widehat{c} - \widehat{c'} - \widehat{ab}) \Longrightarrow \frac{1}{f_{\epsilon_3}(0)\sigma_X^2}\{(2ar_{X2}(\beta_2 + \mu_{\epsilon_2}) - \mu_X, 2a^2 r_{X2} + 1, -2ar_{X2})\mathbf{D}_2(Z_2) - (-\mu_X, 1)D_1(Z_1)\},$$

where $r_{X2} = \sigma_X^2/\sigma_2^2$, $D_1(\cdot)$ is a zero-mean Gaussian process on $\mathbb{R}^2$ with $D_1(\mathbf{0}) = \mathbf{0}$ and

$$E[\{D_1(\mathbf{u}) - D_1(\mathbf{v})\}\{D_1(\mathbf{u}) - D_2(\mathbf{v})\}^{\mathrm{T}}] = f_{\epsilon_3}(0)E\{\mathbf{x}_1\mathbf{x}_1^{\mathrm{T}}|\mathbf{x}_1^{\mathrm{T}}(\mathbf{u}-\mathbf{v})|\} \equiv \Sigma_{D_1}(\mathbf{u}-\mathbf{v}),$$

$\mathbf{D}_2(\cdot)$ is a zero-mean Gaussian process on $\mathbb{R}^3$ with $\mathbf{D}_2(\mathbf{0}) = \mathbf{0}$ and

$$E[\{\mathbf{D}_2(\mathbf{u}) - \mathbf{D}_2(\mathbf{v})\}\{\mathbf{D}_2(\mathbf{u}) - \mathbf{D}_2(\mathbf{v})\}^{\mathrm{T}}] = f_{\epsilon_3}(0)\,E\{\mathbf{x}_2\mathbf{x}_2^{\mathrm{T}}|\mathbf{x}_2^{\mathrm{T}}(\mathbf{u}-\mathbf{v})|\} \equiv \Sigma_{\mathbf{D}_2}(\mathbf{u}-\mathbf{v}),$$

the covariance kernel between $D_1(\cdot)$ and $\mathbf{D}_2(\cdot)$ is

$$f_{\epsilon_3}(0)\,E\{\mathbf{x}_1\mathbf{x}_2^{\mathrm{T}}(|\mathbf{x}_1^{\mathrm{T}}\mathbf{u}| \wedge |\mathbf{x}_2^{\mathrm{T}}\mathbf{v}|)\,1(\mathbf{u}^{\mathrm{T}}\mathbf{x}_1\mathbf{x}_2^{\mathrm{T}}\mathbf{v} > 0)\}$$

with $\wedge$ indicating the minimizer of two real numbers, $(Z_1, Z_2) \sim N(\mathbf{0}, \Sigma)$ with

$$\Sigma = \begin{pmatrix} \Sigma_1 & \Sigma_{12} \\ \Sigma_{12}^{\mathrm{T}} & \Sigma_2 \end{pmatrix}, \quad \Sigma_1 = \frac{E(\mathbf{x}_1\mathbf{x}_1^{\mathrm{T}})^{-1}}{4f_{\epsilon_3}(0)^2},$$

$$\Sigma_{12} = \frac{E(\mathbf{x}_1\mathbf{x}_1^{\mathrm{T}})^{-1}E(\mathbf{x}_1\mathbf{x}_2^{\mathrm{T}})\,E(\mathbf{x}_2\mathbf{x}_2^{\mathrm{T}})^{-1}}{4f_{\epsilon_3}(0)^2}, \quad \Sigma_2 = \frac{E(\mathbf{x}_2\mathbf{x}_2^{\mathrm{T}})^{-1}}{4f_{\epsilon_3}(0)^2},$$

and $(D_1(\cdot), \mathbf{D}_2(\cdot))$ and (Z_1, Z_2) are independent of each other.

We give a few comments on Theorem 2 here. First, when $b = 0$, Condition 1(v) reduces to assume that f_{ϵ_2} and f_{ϵ_3} are continuous at 0, and we strengthen f_{ϵ_3} to be differentiable at 0 for the second-order asymptotic representation (SOAR) of $\widehat{c} - \widehat{c'}$ and $\widehat{b}$. Second, because $\widehat{c} - \widehat{c'}$ and $\widehat{ab}$ have the same asymptotic variance, to compare their efficiency, the second-order expansion is required. From the proof of Theorem 2, both $n^{1/2}(\widehat{c} - \widehat{c'})$ − FOAR and $n^{1/2}\widehat{ab}$ − FOAR are $n^{1/4}$-consistent but have different asymptotic distributions, where FOAR is their common first-order asymptotic representation (revisit (5) to check this fact). These asymptotic distributions are uncorrelated with (although not independent of) the first-order asymptotic distributions $(Z_1^{\mathrm{T}}, Z_2^{\mathrm{T}})$ of $((\widehat{\beta}_1, \widehat{c}), (\widehat{\beta}_3, \widehat{c'}, \widehat{b}))$. This is dramatically different from the SOAR in LS mediation analysis. From typical Edgeworth expansions, we know that the SOAR of $\widehat{c} - \widehat{c'}$ (and $\widehat{ab}$, which is the same as $\widehat{c} - \widehat{c'}$) is $n^{1/2}$-consistent rather than $n^{1/4}$-consistent. Third, the asymptotic distribution of $n^{3/4}(\widehat{c} - \widehat{c'} - \widehat{ab})$ follows a (variance) mixture normal distribution, whose density is tedious to express in the explicit form. We will provide an explicit formula for it when $a = 0$ in Section 3.4. Fourth, when $a = 0$, $\sigma_C^2 = 0$, so the asymptotic distributions of $\widehat{c} - \widehat{c'}$ and $\widehat{ab}$ will degenerate and further refinements are required; see Section 3.4. Fifth, from our simulations in Section 4, although the FOAR of $\widehat{c} - \widehat{c'}$ and $\widehat{ab}$ are the same when $b = 0$, the variance of the SOAR of $\widehat{c} - \widehat{c'}$ is much larger than that of $\widehat{ab}$ when a is relatively small. In other words, $\widehat{ab}$ is still more efficient than $\widehat{c} - \widehat{c'}$ in finite samples as in Case one.

3.4. *Case three: $a = b = 0$*

We finally consider the case with $a = b = 0$ in the following Theorem 3.

THEOREM 3. *When $a = b = 0$, if Condition 1 holds, then*

$$n\widehat{ab} \Longrightarrow \frac{1}{4 f_{\epsilon_2}(0) f_{\epsilon_3}(0) \sigma_X \sigma_2} z_1 z_2,$$

and if further assume Condition 2 holds, then

$$n^{3/4}(\widehat{c} - \widehat{c'}) \Longrightarrow \frac{(-\mu_X, 1)}{f_{\epsilon_3}(0)\sigma_X^2} \left\{ D_2(Z_2) - D_1(Z_1) \right\},$$
$$n^{3/4}(\widehat{c} - \widehat{c'} - \widehat{ab}) \Longrightarrow \frac{(-\mu_X, 1)}{f_{\epsilon_3}(0)\sigma_X^2} \left\{ D_2(Z_2) - D_1(Z_1) \right\},$$

where z_1 and z_2 are two independent standard normal random variables, $D_1(\cdot)$, Z_1 and Z_2 are defined in Theorem 2, $D_2(\cdot)$ is the first two elements of $\mathbf{D}_2(\cdot)$ in Theorem 2 so is a zero-mean Gaussian process on $\mathbb{R}^2$ with $D_2(\mathbf{0}) = \mathbf{0}$ and

$$E\left[\left\{D_2(\mathbf{u}) - D_2(\mathbf{v})\right\}\left(D_2(\mathbf{u}) - D_2(\mathbf{v})\right)^{\mathrm{T}}\right] = f_{\epsilon_3}(0) E\left\{\mathbf{x}_1\mathbf{x}_1^{\mathrm{T}} \left|\mathbf{x}_2^{\mathrm{T}}(\mathbf{u} - \mathbf{v})\right|\right\} \equiv \Sigma_{D_2}(\mathbf{u} - \mathbf{v}),$$

the covariance kernel between $D_1(\cdot)$ and $D_2(\cdot)$ is

$$f_{\epsilon_3}(0) E\left\{\mathbf{x}_1\mathbf{x}_1^{\mathrm{T}} (|\mathbf{x}_1^{\mathrm{T}}\mathbf{u}| \wedge |\mathbf{x}_2^{\mathrm{T}}\mathbf{v}|) 1(\mathbf{u}^{\mathrm{T}}\mathbf{x}_1\mathbf{x}_2^{\mathrm{T}}\mathbf{v} > 0)\right\} \equiv \Sigma_{D_{12}}(\mathbf{u} - \mathbf{v}),$$

and $(D_1(\cdot), D_2(\cdot))$ and (Z_1, Z_2) are independent of each other.

We give a few comments on Theorem 3 here. First, it is interesting to observe that $\widehat{ab}$ and $\widehat{c} - \widehat{c'}$ have different convergence rates. This is dramatically different from Cases one and two where they have the same convergence rates $n^{1/2}$. This is also very different from the case in LS mediation analysis where $\widehat{ab} = \widehat{c} - \widehat{c'}$ and both are n-consistent. Since $\widehat{ab}$ has a faster convergence rate than $\widehat{c} - \widehat{c'}$, it is superior, same as in Cases One and Two; also, it implies $n^{3/4}(\widehat{c} - \widehat{c'} - \widehat{ab})$ and $n^{3/4}(\widehat{c} - \widehat{c'})$ have the same asymptotic distribution. Second, when $a = 0$, $\left\{2ar_{X2}(\beta_2 + \mu_{\epsilon_2}) - \mu_X, 2a^2 r_{X2} + 1, -2ar_{X2}\right\} \mathbf{D}_2(Z_2)$ in the asymptotic distribution of $\widehat{c} - \widehat{c'} - \widehat{ab}$ in Theorem 2 will reduce to $(-\mu_X, 1, 0)\,\mathbf{D}_2(Z_2) = (-\mu_X, 1)\, D_2(Z_2)$ as indicated in Theorem 3. Third, when $a = b = 0$, Σ_1, Σ_{12} and Σ_2 in Theorem 2 can be simplified such that we can write $Z_2^{\mathrm{T}} = (Z_1^{\mathrm{T}} + (z_3, 0), z_2)$ with (z_2, z_3) independent of Z_1, $Z_1 \sim N(\mathbf{0}, \Sigma_1)$, and

$$(z_2, z_3)^{\mathrm{T}} \sim N\left(0, \frac{1}{4 f_{\epsilon_3}(0)^2 \sigma_2^2}\begin{pmatrix} 1 & -\mu_M \\ -\mu_M & \mu_M^2 \end{pmatrix}\right);$$

see Appendix F for details. Fourth, $(-\mu_X, 1)\{D_2(Z_2) - D_1(Z_1)\}/\{\sigma_X^2 f_{\epsilon_3}(0)\}$ follows a (variance) mixture normal distribution. Its density is

$$g(x) = \int f(x \mid 0, \sigma_D^2(\mathbf{z}_1, \mathbf{z}_2)) f(\mathbf{z}_1, \mathbf{z}_2 \mid \mathbf{0}, \Sigma) d\mathbf{z}_1 d\mathbf{z}_2,$$

where $f(\mathbf{x} \mid \mathbf{0}, \Omega)$ is the pdf of the normal distribution with mean $\mathbf{0}$ and variance matrix Ω, and

$$\sigma_D^2(\mathbf{z}_1, \mathbf{z}_2) = \frac{(-\mu_X, 1)\left\{\Sigma_{D_1}(\mathbf{z}_1) + \Sigma_{D_2}(\mathbf{z}_2) - 2\Sigma_{D_{12}}(\mathbf{z}_1, \mathbf{z}_2)\right\}(-\mu_X, 1)^{\mathrm{T}}}{\sigma_X^4 f_{\epsilon_3}(0)^2}$$

with $\mathbf{z}_1, \mathbf{z}_2 \in \mathbb{R}^2$. The asymptotic variance of $n^{3/4}(\widehat{c} - \widehat{c'})$ is

$$\frac{E\left[(-\mu_X, 1)\left\{\Sigma_{D_1}(Z_1) + \Sigma_{D_2}(Z_2) - 2\Sigma_{D_{12}}(Z_1, Z_2)\right\}(-\mu_X, 1)^{\mathrm{T}}\right]}{\sigma_X^4 f_{\epsilon_3}(0)^2}.$$

When $\mu_X = 0$, it can be further simplified; see Appendix F for details. Fifth, from Theorem 3,

$$\mathrm{var}(\widehat{ab}) \approx \frac{1}{16n^2}\frac{1}{f_{\epsilon_2}(0)^2 f_{\epsilon_3}(0)^2}\frac{1}{\sigma_X^2\sigma_2^2}E\left(z_1^2 z_2^2\right) = \frac{1}{16n^2}\frac{1}{f_{\epsilon_2}(0)^2 f_{\epsilon_3}(0)^2}\frac{1}{\sigma_X^2\sigma_2^2}$$

which is equal to $\pi^2/(4n^2\sigma_X^2)$ when ϵ_2 and ϵ_3 follow the standard normal, where $\approx$ means higher order terms are omitted throughout this paper.

4. SIMULATIONS

4.1. *Simulation designs*

Because β_2, β_3 and c' will not affect the asymptotic distribution of any parameter of interest as indicated in Theorems 1, 2 and 3, we set $\beta_2 = \beta_3 = 0$ and $c' = 1$ throughout our simulations. The sample size is set at $n = 200, 500, 1000, 2000, 5000, 10000$, and the replication number is set at $N = 10000$. To save space, we report only the simulation results for $n = 200$ and 1000 in our three tables as the results for the other sample sizes are consistent with these two sample sizes. Anyway, in our four figures, we use information from all sample sizes. In the three tables, we report the finite-sample MSE companioned with the MSE predicted by the asymptotic theory for the LS and two LAD estimators. As the bias is ignorable in both finite samples and large samples, the MSE is roughly equal to the variance. To satisfy Conditions 1 and 2, we set $X \sim N(0, 1)$, $\epsilon_2 \sim N(0, 1)$, and ϵ_3 follows three popular distributions: (I) $N(0, 1)$, (II) $0.9N(0, 1) + 0.1N(0, 10^2)$, and (III) t_2.

4.2. *Case one: $b \neq 0$*

When $b \neq 0$, we set $a = b = 0.14, 1.4$, corresponding to small and large mediation effects. From Table 1, we can draw the following conclusions. First, the large-sample MSE matches the finite-sample MSE very well, which implies that the convergence rates of all three estimators are $n^{1/2}$ as predicted by Theorem 1. Second, the product LAD estimator is the most efficient except in case (I) where the LS estimator is the most efficient. Third, the difference estimator is less efficient than the product estimator and the efficiency of the former gets closer to that of the

Table 1. *MSE ($\times 10^{-3}$) for LS and two LAD estimates*

		$a=b=0.14$			$a=b=1.4$		
ϵ_3	n	MSE_{LS}	MSE_{P}	MSE_{D}	MSE_{LS}	MSE_{P}	MSE_{D}
(I)	200	0.22	0.38	1.69	19.68	31.03	36.46
		0.20	0.31	1.56	19.60	30.79	35.82
	1000	0.04	0.06	0.32	3.94	6.20	7.17
		0.04	0.06	0.31	3.92	6.16	7.16
(II)	200	1.44	0.42	1.86	118.82	34.35	41.69
		1.17	0.33	1.69	116.62	33.18	40.36
	1000	0.25	0.07	0.35	23.47	6.57	8.17
		0.23	0.07	0.34	23.32	6.64	8.07
(III)	200	1.93	0.44	2.00	149.37	35.50	44.11
		∞	0.35	1.79	∞	35.00	43.73
	1000	0.68	0.07	0.36	32.18	7.05	8.75
		∞	0.07	0.36	∞	7.00	8.73

For each ϵ_3 distribution and n, the first row is the finite-sample MSE and the second row is the MSE predicted by the asymptotic theory.

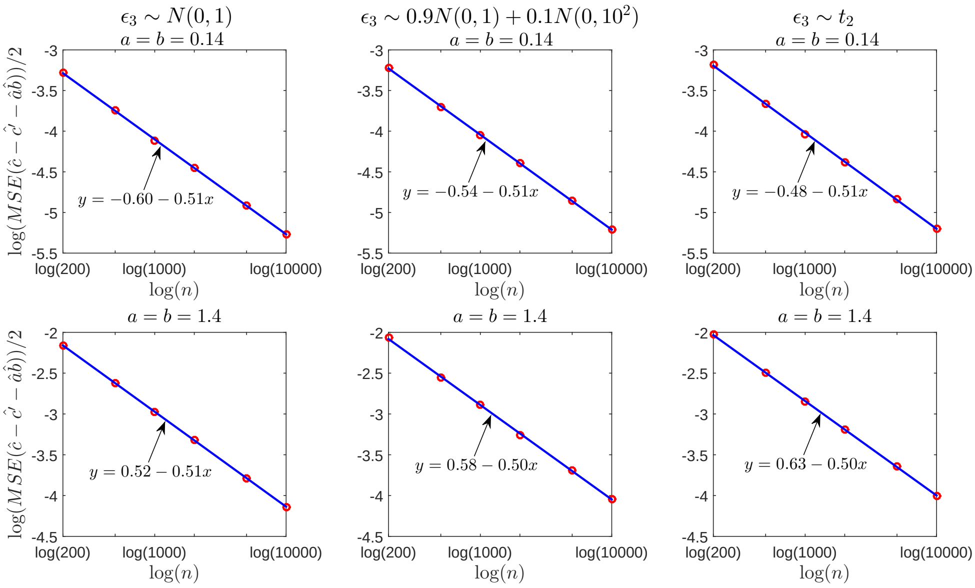

Fig. 1. $\log\{\mathrm{MSE}(\widehat{c}-\widehat{c'}-\widehat{a}\widehat{b})\}/2$ against $\log n$ for two (a,b) values and three ϵ_3 distributions.

latter when the mediation effect gets larger. Finally, the MSE is larger when ϵ_3 has a heavier tail (or $f_{\epsilon_3}(0)$ is smaller) for all three estimators.

Figure 1 shows $\log\{\mathrm{MSE}(\widehat{c}-\widehat{c'}-\widehat{a}\widehat{b})\}/2$ against $\log n$ when $b \neq 0$; it indicates that the convergence rate of $\widehat{c}-\widehat{c'}-\widehat{a}\widehat{b}$ is indeed $n^{1/2}$ when $b \neq 0$ as predicted by Theorem 1. Furthermore, since the asymptotic variance of $\widehat{c}-\widehat{c'}-\widehat{a}\widehat{b}$ is roughly $e^{2\hat{\theta}_0}$, where $\hat{\theta}_0$ is the intercept of regressing $\log\{\mathrm{MSE}(\widehat{c}-\widehat{c'}-\widehat{a}\widehat{b})\}/2$ on $\log n$, Figure 1 indicates that this asymptotic variance increases with the heaviness of ϵ_3's tail and the magnitude of the mediation effect. For example, if $a=0.14$, the ratio of the asymptotic variances when $\epsilon_3 \sim t_2$ and $N(0,1)$ is $e^{-2\times 0.48}/e^{-2\times 0.60} \approx 1.29$, and if $\epsilon_3 \sim N(0,1)$, this ratio when $a=1.4$ and $a=0.14$ is $e^{2\times 0.52}/e^{-2\times 0.60} \approx 9.36$.

Table 2. *MSE ($\times 10^{-3}$) for LS and two LAD estimates*

		$b=0,a=0.14$			$b=0,a=1.4$		
ϵ_3	n	MSE$_{LS}$	MSE$_P$	MSE$_D$	MSE$_{LS}$	MSE$_P$	MSE$_D$
(I)	200	0.13	0.23	0.92	9.96	15.86	16.68
		0.10	0.15	0.15	9.80	15.39	15.39
	1000	0.02	0.03	0.10	1.95	3.06	3.11
		0.02	0.03	0.03	1.96	3.08	3.08
(II)	200	1.38	0.27	1.09	112.23	18.81	19.74
		1.07	0.18	0.18	106.82	17.79	17.79
	1000	0.22	0.04	0.12	21.56	3.71	3.78
		0.21	0.04	0.04	21.36	3.56	3.56
(III)	200	1.43	0.29	1.15	169.12	20.04	20.92
		∞	0.20	0.20	∞	19.60	19.60
	1000	0.34	0.04	0.12	39.06	3.98	4.04
		∞	0.04	0.04	∞	3.92	3.92

4.3. *Case two: $b=0$ but $a\neq 0$*

When $b=0$, we set $a=0.14,1.4$. From Table 2, we can draw the following conclusions. First, the large-sample MSE matches the finite-sample MSE very well except for the difference estimator when $a=0.14$; as a result, the finite-sample MSEs of the two LAD estimators are very different when $a=0.14$ even if their asymptotic MSEs are the same as predicted by Theorem 2. When $a=0.14$, the FOAR is not a good approximation to the asymptotic distribution of $\widehat{c}-\widehat{c'}$. Actually, the SOAR takes in charge. Figure 2 shows $\log\{\text{MSE}(\widehat{c}-\widehat{c'})-\sigma_C^2/n\}/2$ against $\log n$, where $\text{MSE}(\widehat{c}-\widehat{c'})-\sigma_C^2/n$ is approximately the variance of the SOAR by recalling that the FOAR and the SOAR are asymptotically uncorrelated. As mentioned after Theorem 2, the root mean square error (RMSE) of the SOAR in the LAD estimates is $O(n^{-3/4})$ rather than $O(n^{-1})$ as in the LS estimate. Because the asymptotic variance of the SOAR is large for $\widehat{c}-\widehat{c'}$ when a is small, its effect on the MSE cannot be neglected; on the other hand, the counterpart for $\widehat{ab}$ is very small, so the asymptotic variance of the FOAR is a good approximation to the MSE. For example, when $\epsilon_3\sim N(0,1)$ and $a=0.14$, the asymptotic variance of the SOAR of $\widehat{c}-\widehat{c'}$ is $e^{2\times 0.36}\approx 2.06$, while that of $\widehat{ab}$ is close to zero. From Fig. 2, we can see that the asymptotic variance of the SOAR of $\widehat{c}-\widehat{c'}$ also increases with the heaviness of ϵ_3's tail, just as the asymptotic variance of the FOAR as indicated Table 2. Second, the second and third conclusions in Case one still hold.

As Fig. 1, Figure 3 shows $\log\{\text{MSE}(\widehat{c}-\widehat{c'}-\widehat{ab})\}/2$ against $\log n$ when $b=0$ but $a\neq 0$. Different from Fig. 1, Figure 3 indicates that the convergence rate of $\widehat{c}-\widehat{c'}-\widehat{ab}$ is $n^{3/4}$ when $b=0$ (rather than $n^{1/2}$ when $b\neq 0$), which matches the prediction of Theorem 2. Also, the asymptotic variance of $\widehat{c}-\widehat{c'}-\widehat{ab}$ increases with the heaviness of ϵ_3's tail, but does not seem to increase with a as in the $b\neq 0$ case. Comparing Fig. 2 and Fig. 3, we can see that $\text{MSE}(\widehat{c}-\widehat{c'}-\widehat{ab})$ and $\text{MSE}(\widehat{c}-\widehat{c'})-\sigma_C^2/n$ when $a=0.14$ are quite close, which is because the asymptotic variance of the SOAR of $\widehat{ab}$ is close to zero so that both MSEs are roughly the variance of the SOAR of $\widehat{c}-\widehat{c'}$.

4.4. *Case three: $a=b=0$*

The first, second and fourth conclusions from Table 1 still apply to Table 3; especially, the first conclusion implies that the convergence rates of the LS and product estimators are n while the convergence rate of the difference estimator is $n^{3/4}$ (see Appendix F for more discussions), which is in accordance with Theorem 3 and dramatically different from Cases one and two where

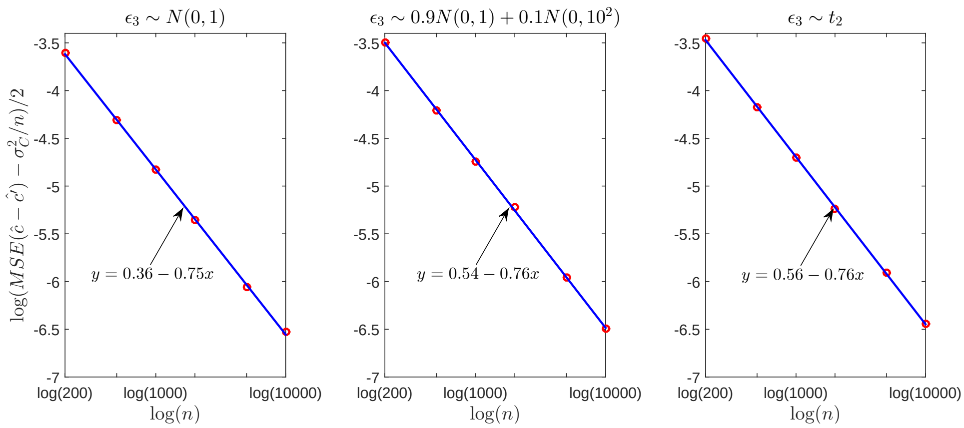

Fig. 2. $\log\{\text{MSE}(\widehat{c}-\widehat{c'})-\sigma_C^2/n\}/2$ against $\log n$ for $a=0.14$, $b=0$ and three ϵ_3 distributions

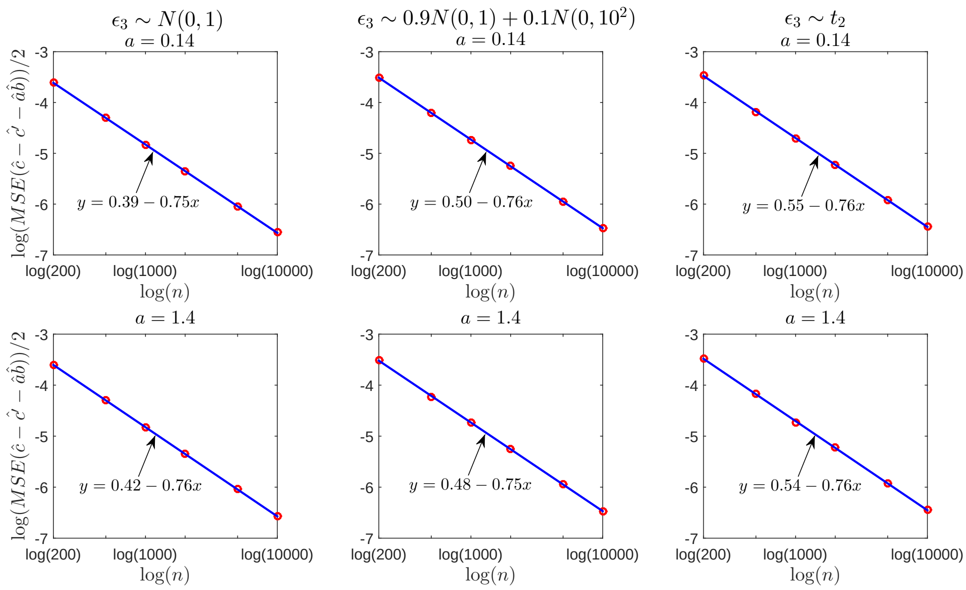

Fig. 3. $\log\{\text{MSE}(\widehat{c}-\widehat{c'}-\widehat{a}\widehat{b})\}/2$ against $\log n$ for $b=0$, two a values and three ϵ_3 distributions.

Table 3. *MSE* ($\times 10^{-5}$) *for LS and two LAD estimates when* $a=b=0$

ϵ_3	n	MSE_{LS}	MSE_P	MSE_D
(I)	200	2.59	6.26	72.40
		2.50	6.17	70.06
	1000	0.10	0.25	6.32
		1.10	0.25	6.27
(II)	200	27.77	7.99	92.50
		27.25	7.13	80.96
	1000	1.11	0.29	7.47
		1.09	0.29	7.24
(III)	200	33.91	8.58	99.62
		∞	7.85	89.21
	1000	1.44	0.32	8.42
		∞	0.31	7.98

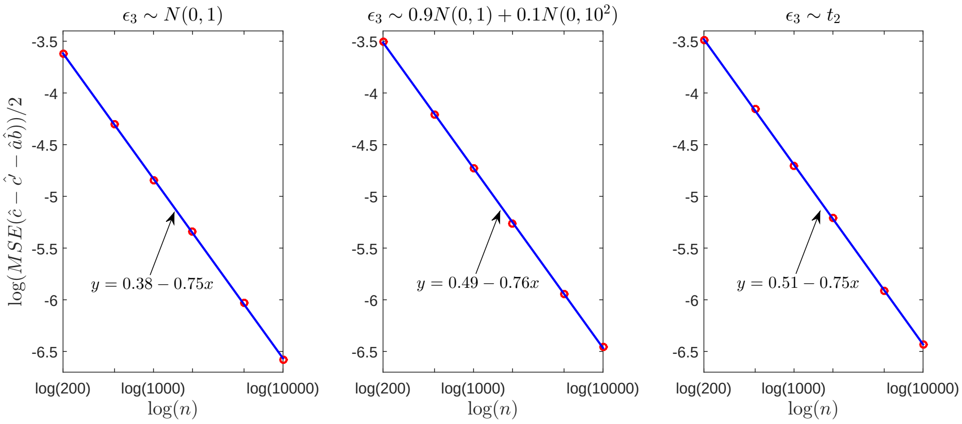

Fig. 4. $\log\{MSE(\widehat{c} - \widehat{c'} - \widehat{a}\widehat{b})\}/2$ against $\log n$ for $a = b = 0$ and three ϵ_3 distributions.

all three estimators are $n^{1/2}$-consistent. As Fig. 1 and Fig. 3, Figure 4 shows $\log\{\mathrm{MSE}(\widehat{c} - \widehat{c'} - \widehat{a}\widehat{b})\}/2$ against $\log n$ when $a = b = 0$. As predicted by Theorem 3, the convergence rate of $\widehat{c} - \widehat{c'} - \widehat{a}\widehat{b}$ is $n^{3/4}$, same as in Case two. Also, similar to Cases One and Two, the asymptotic variance of $\widehat{c} - \widehat{c'} - \widehat{a}\widehat{b}$ increases with the heaviness of ϵ_3's tail.

5. DISCUSSION

This paper develops asymptotic theories for two forms of mediation effect estimates and shows their asymptotic nonequivalence in a basic LAD mediation model. The LAD method can be easily generalized to multilevel mediation models (Hox, 2002; Preacher et al., 2010) and multi-mediator models (VanderWeele & Vansteelandt, 2014). In order to further improve the estimation efficiency, many other robust methods can be applied to the mediation model, including the weighted quantile average regression (Zhao & Xiao, 2014), the differenced method (Wang et al., 2019), and the general M-estimation method (Huber & Ronchetti, 2009). Analyses in these general models and for other robust methods are left for future research, but the results in this paper will definitely shed some lights on these extensions.

ACKNOWLEDGEMENT

We thank Professor Keith Knight for providing us his papers on second-order asymptotics of quantile regression.

SUPPLEMENTARY MATERIAL

Supplementary material available online includes the proofs of Proposition 1, Theorems 1-3, Corollary 1, and simplifications in Case three and simulations.

REFERENCES

ANGRIST, J., CHERNOZHUKOV, V. & FERNÁNDEZ-VAL, I. (2006). Quantile regression under misspecification, with an application to the U.S. wage structure. *Econometrica* **74**, 539–563.

BARON, R. M. & KENNY, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *J. Pers. Soc. Psychol.* **51**, 1173–1182.

Bind, M.-A., VanderWeele, T. J., Schwartz, J. D. & Coull, B. A. (2017). Quantile causal mediation analysis allowing longitudinal data. *Stat. Med.* **36**, 4182–4195.

Field, A. P. & Wilcox, R. R. (2017). Robust statistical methods: A primer for clinical psychology and experimental psychopathology researchers. *Behav. Res. Ther.* **98**, 19–38.

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

Hox, J. J. (2002). *Multilevel Analysis: Techniques and Applications*. Mahwah, NJ: Erlbau.

Huber, P. J. & Ronchetti, E. M. (2009). *Robust Statistics*. New York: Wiley, 2nd ed.

Knight, K. (1998). Limiting distributions for L_1 regression estimators under general conditions. *Ann. Statist.* **26**, 755–770.

Koenker, R. (2005). *Quantile Regression*. New York: Cambridge University Press.

Lindquist, M. A. (2012). Functional causal mediation analysis with an application to brain connectivity. *J. Am. Stat. Assoc.* **107**, 1297–1309.

MacKinnon, D. P. (2008). *Introduction to Statistical Mediation Analysis*. New York.

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G. & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychol. Methods* **7**, 83–104.

MacKinnon, D. P., Warsi, G. & Dwyer, J. H. (1995). A simulation study of mediated effect measures. *Multivar. Behav. Res.* **30**, 41–62.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychol. Bull.* **105**, 156–166.

Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika* **82**, 669–710.

Pearl, J. (2009). *Causality: Models, Reasoning, and Inference*. New York: Cambridge University Press, 2nd ed.

Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econ. Theory* **7**, 186–199.

Preacher, K. J. (2015). Advances in mediation analysis: A survey and synthesis of new developments. *Annu. Rev. Psychol.* **66**, 825–852.

Preacher, K. J., Zyphur, M. J. & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychol. Methods* **15**, 209–233.

Shen, E., Chou, C.-P., Pentz, M. A. & Berhane, K. (2014). Quantile mediation models: A comparison of methods for assessing mediation across the outcome distribution. *Multivar. Behav. Res.* **49**, 471–485.

VanderWeele, T. J. (2015). *Explanation in Causal Inference: Methods for Mediation and Interaction*. Oxford, UK: Oxford University Press.

VanderWeele, T. J., Hong, G. L., Jones, S. M. & Brown, J. L. (2013). Mediation and spillover effects in group-randomized trials: A case study of the 4Rs educational intervention. *J. Am. stat. Assoc.* **108**, 469–482.

VanderWeele, T. J. & Vansteelandt, S. (2014). Mediation analysis with multiple mediators. *Epidemiol. Methods* **2**, 95–115.

Wang, W. W., Yu, P., Lin, L. & Tong, T. (2019). Robust estimation of derivatives using locally weighted least absolute deviation regression. *J. Mach. Learn. Res.* **60**, 1–49.

Wang, W. W., Yu, P., Zhou, Y., Tong, T. & Liu, Z. (2020). Equivalence of two least-squares estimators for indirect effects. *Submit to Curr. Psychol.* , under review.

Wilcox, R. R. (2017). *Introduction to Robust Estimation & Hypothesis Testing*. Amsterdam: Elsevier, 4th ed.

Yuan, Y. & MacKinnon, D. P. (2014). Robust mediation analysis based on median regression. *Psychol. Methods* **19**, 1–20.

Zhao, Z. & Xiao, Z. (2014). Efficient regression via optimally combining quantile information. *Econ. Theory* **30**, 1272–1314.

Zu, J. & Yuan, Y. (2010). Local influence and robust procedures for mediation analysis. *Multivar. Behav. Res.* **45**, 1–44.