Statistica Sinica **31** (2021), 1-23
doi:https://doi.org/10.5705/ss.202019.0150

TOTAL-EFFECT TEST IS SUPERFLUOUS FOR ESTABLISHING COMPLEMENTARY MEDIATION

Yingkai Jiang[1], Xinshu Zhao[2], Lixing Zhu[3], Jun S. Liu[4] and Ke Deng[1]

[1] *Tsinghua University,* [2] *University of Macau,*
[3] *Hong Kong Baptist University and* [4] *Harvard University*

Abstract: Mediation, which means that an independent variable X affects a dependent variable Y through a mediator M, is a key concept in causal inference. For establishing mediation, there is a long debate on whether to require the "total effect" of X on Y to be statistically significant. It has been shown that total-effect test can erroneously reject "competitive mediation". For "complementary mediation", however, the situation becomes more complicated. This article provides an explicit proof that the total effect is statistically significant whenever mediated effect and direct effect bear the same sign and are both significant, as long as the least square estimation (LSE) and F-tests are used to estimate and test mediation effects. We also show that the similar result can be obtained when the Sobel test is used. Our results support the growing agreement that total-effect test is unnecessary for establishing any type of mediation.

Key words and phrases: Bp, complementary mediation, hypothesis testing, linear model, mediation analysis, percentage coefficient, percentage scale, total-effect test.

1. Introduction

The concept of *mediation* has been broadly used in many areas of social sciences, which generally means that an independent variable X affects a dependent variable Y through a *mediator* M. It plays an important role in understanding causal mechanism, and is the focus of many research projects. The classic mediation model (Baron and Kenny (1986)) can be represented by the linear regression below:

$$M = i_M + aX + \varepsilon_M, \tag{1.1}$$

$$Y = i_Y + bM + dX + \varepsilon_Y, \tag{1.2}$$

where the errors are assumed to follow independent normal distributions

$$\varepsilon_M \sim N(0, \sigma_M^2), \quad \varepsilon_Y \sim N(0, \sigma_Y^2).$$

Corresponding author: Ke Deng, Center for Statistical Science & Department of Industry Engineering, Tsinghua University, Haidian, Beijing, China. E-mail: kdeng@tsinghua.edu.cn.

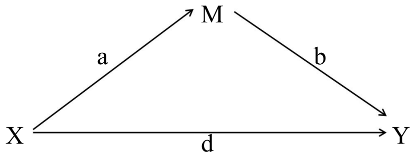

Figure 1. Mediation. Adapted from Baron and Kenny (1986), Figure 3.

There are two paths from X to Y: a *direct path* "$X \rightarrow Y \mid M$" representing the *direct effect* of X on Y while M is controlled (which equals to d), and an *indirect path* "$X \rightarrow M \rightarrow Y$" representing the *mediated effect* of X on Y via the mediator M (which equals to $a \times b$). Reorganizing (1.1) and (1.2), we obtain the following linear model with X and Y only:

$$Y = i_Y^* + cX + \varepsilon_Y^*, \tag{1.3}$$

where $i_Y^* = i_Y + bi_M$, $\varepsilon_Y^* = \varepsilon_Y + b\varepsilon_M$, and $c = a \times b + d$ stands for the *total effect* of X on Y combining the indirect effect $a \times b$ and the direct effect d. Based on the relationship between the direct effect d and the mediated effect $a \times b$, mediation via the mediator M can be classified into three types according to Zhao, Lynch and Chen (2010): the *competitive mediation*, which happens when the direct and indirect paths bear opposite signs so that their effects offset each other; the *complementary mediation*, which happens when the direct and indirect paths bear the same sign so that their effects complement each other; and the *indirect-only mediation*, which happens when the direct effect $d = 0$ while the indirect effect $a \times b \neq 0$.

It would be straightforward to establish mediation if the parameters are precisely known (e.g., mediation exists in the classic mediation model if $a \times b \neq 0$). The task becomes challenging in practice as typically only data instead of parameters are available. Baron and Kenny's classic procedure (Baron and Kenny (1986)) requires the simple correlation between X and Y to be significant, in addition to the significance of the indirect effect $a \times b$. MacKinnon, Warsi and Dwyer (1995) demonstrated that the simple correlation is exactly the *total effect* $c = a \times b + d$ under the classic model of mediation. Therefore, Baron and Kenny's classic procedure requires both the *indirect-effect test* for $a \times b$ and the *total-effect test* for c to be significant. While many researchers (Judd and Kenny (1981); Rose et al. (2004); Mathieu and Taylor (2006)) followed Baron and Kenny (1986) to require the *total-effect test* for establishing any mediation, some (Collins, Graham and Flaherty (1998); Kenny, Kashy and Bolger (1998); Rose et al. (2000);

MacKinnon, Krull and Lockwood (2000); MacKinnon et al. (2002); Shrout and Bolger (2002)) recommended suspending the test for some types of mediation, leading to a long debate among researchers, especially social scientists.

An obvious argument against the total-effect test regards competitive mediation: when the direct and indirect paths bear opposite signs, their effects offset each other, and hence the total effect can be non-significant even when the mediated path is strong. The phenomenon is well-known, as shown in the long list of publications on the topic (Conger (1974); Velicer (1978); McFatter (1979); Davis (1985); Hamilton (1987); Cohen (1988); Tzelgov and Henik (1991); Kenny, Kashy and Bolger (1998); MacKinnon, Krull and Lockwood (2000); Shrout and Bolger (2002); Lord and Novick (2008); Hayes (2009); Zhao, Lynch and Chen (2010); Rucker et al. (2011)). There are also simulated data (McFatter (1979); Collins, Graham and Flaherty (1998); Hayes (2009)) and real-data examples (Zhao (1997); Zhao, Lynch and Chen (2010); Li et al. (2013)) in support of the argument. A second argument, offered by Shrout and Bolger (2002), is that when the independent variable X occurs temporally long before the dependent variable Y, or when the expected effect size is small, it would be too difficult for the mediated effect to survive the total-effect test. The authors' hypothetical example was how out-of-home placement of children affects their substance abuse during adulthood. A third argument, by Zhao, Chen and Tong (2011), is that in an *indirect-only mediation* where the mediated $a \times b$ path is significant but the direct d path is not, the large statistical error of d path relative to its effect size may inflate the statistical error of the total effect c relative to its effect size. A total-effect test in this situation may produce a misleading non-significant c when mediation $a \times b$ is in effect strong. There is also a real data example (Zhao et al. (1994)) in support of the argument. A fourth argument, which Zhao, Lynch and Chen (2010) mentioned in passing, is that in a *complementary mediation*, where the direct and indirect paths bear the same sign and both are significant, the total-effect test always passes, making the test superfluous. Encouraged by these arguments, more recent authors (Hayes (2009); MacKinnon and Fairchild (2009); Zhao, Lynch and Chen (2010); Rucker et al. (2011); Zhao, Chen and Tong (2011)) advocated waiving the test for establishing any type of mediation.

While the experts seem to agree to suspend the total-effect test for competitive mediation, there are less agreements on whether to require the test for the other two types of mediation, especially complementary mediation. The detailed debate can be found in Shrout and Bolger (2002), Rose et al. (2004), Wen et al. (2004), Mathieu and Taylor (2006), and Wen and Ye (2014). Although various arguments have been advanced for all sides, no explicit statistical formulation or solid mathematical proof for the original problem is available even for the classic mediation model.

This article addresses this issue. By reformulating the issue into a geometric problem about the rejection regions of the tests involved, we provide an explicit proof that the total effect is statistically significant whenever mediated and direct effects bear the same sign and are both significant, as long as the least square estimation (LSE) and the F-test or the Sobel test (Sobel (1982)) are used to estimate and test mediation effects. Considering that the LSE-F and LSE-Sobel frameworks are the classic standard approaches for mediation analysis, our finding provides support to the growing agreement that the total-effect test is unnecessary for establishing any type of mediation.

2. Frameworks to Establish Complementary Mediation

In classic mediation model, treating the direct effect d and the mediation effect $a \times b$ as unknown constants, we obtain the obvious equivalence between "$c = a \times b + d = 0$" and "$a \times b = 0$ and $d = 0$" as long as $a \times b$ and d bear the same sign. Furthermore, it seems intuitive that the same would hold for statistical inference, i.e., if the two paths d and $a \times b$ bear the same sign and are both significant, their combination $c = a \times b + d$ must point in the same direction and also be statistically significant. If the intuition is correct, we would be able to assure the significance of total effect c by testing the significance of a, b, and d, and the total effect test would be redundant. To the best of our knowledge, however, there is no explicit theoretical proof for this intuition, perhaps due to the complexity of the statistical inference. The lack of theoretical guarantee has contributed to confusions, disagreements, and continued debates on the role of total effect test for establishing mediation.

There are different ways of estimating mediation. Baron and Kenny (1986) suggested to estimate (a, b, d, c) by their LSEs $(\hat{a}, \hat{b}, \hat{d}, \hat{c})$ and claim the indirect path of mediation effect by the Sobel test, which tests

$$H_0: \ a \times b = 0 \quad \text{vs} \quad H_1 : a \times b \neq 0 \tag{2.1}$$

with statistic

$$S = \frac{\hat{a}\hat{b}}{\left(\hat{a}^2 \mathrm{Var}(\hat{b}) + \hat{b}^2 \mathrm{Var}(\hat{a})\right)^{1/2}},$$

whose asymptotic distribution under the null is the standard normal. The LSE-Sobel framework enjoys the advantage of straightforward intuition as it infers the indirect mediation effect $a \times b$ directly with a single test. Its limitation lies in the fact that it is not an exact test as the distribution of the test statistic S depends on the values of a and b.

Alternatively, Judd and Kenny (1981) suggested to establish mediation by estimating and testing a, b, d, c separately, given that the original test in (2.1) can be recast as an equivalent problem below:

$$H_0: \; a = 0 \text{ or } b = 0 \quad \text{vs} \quad H_1: a \neq 0 \text{ and } b \neq 0, \tag{2.2}$$

which can be resolved by checking whether $a \neq 0$ and $b \neq 0$ separately via the tests below:

$$H_0: \; a = 0 \quad \text{vs} \quad H_1: a \neq 0, \tag{2.3}$$

$$H_0: \; b = 0 \quad \text{vs} \quad H_1: b \neq 0. \tag{2.4}$$

If the null hypothesis is rejected for both (2.3) and (2.4), it appears that the null hypothesis for test (2.2) should be rejected too. A natural way to implement this idea is the LSE-F framework, in which a, b, d, c are estimated by LSE and tested by the F-test. Because the F-tests for a, b, d, c are all exact, the LSE-F framework enjoys the theoretical convenience that the LSE-Sobel framework does not.

Moreover, to deal with cases where the noise terms ε_M and ε_Y follow a heavy-tail distribution, e.g., Laplace distribution, Pollard (1991) proposed the LAD-Z framework, which follows a similar strategy as the LSE-F framework. More precisely, in LAD-Z one estimates the regression coefficients by the more robust *least absolute deviation estimation* (LAD) and tests their significance by the Z-test: comparing the Z-statistic $z_j = |\check{\beta}_j|/\mathrm{sd}(\check{\beta}_j)$ with the standard normal distribution to establish the statistical significance, where $\check{\beta}_j$ is the LAD estimate of regression coefficient β and $\mathrm{sd}(\check{\beta}_j)$ is the estimated standard deviation of $\check{\beta}_j$. MacKinnon, Warsi and Dwyer (1995) provided a comprehensive review of the different frameworks and compared their performance via simulations.

3. Main Results

3.1. The major theorem

This study focuses on LSE-F framework. Let $\hat{a}$, $(\hat{b}, \hat{d})$, and $\hat{c}$ be the LSEs of the coefficients a, (b, d), and c in regression models (1.1), (1.2), and (1.3), respectively. We use $\mathcal{R}_a(\alpha)$, $\mathcal{R}_b(\alpha)$, $\mathcal{R}_d(\alpha)$ and $\mathcal{R}_c(\alpha)$ to denote the rejection regions of the corresponding F-tests under the critical level $\alpha \in (0, 1)$, and p_a, p_b, p_d and p_c are the corresponding p-values. We note that the question of "whether the total-effect test is superfluous for establishing complementary mediation" can be addressed by verifying whether $\mathcal{R}_a(\alpha) \cap \mathcal{R}_b(\alpha) \cap \mathcal{R}_d(\alpha) \subseteq \mathcal{R}_c(\alpha)$ for all $\alpha \in (0,1)$. Apparently, if $\mathcal{R}_a(\alpha) \cap \mathcal{R}_b(\alpha) \cap \mathcal{R}_d(\alpha)$ is always a subset of $\mathcal{R}_c(\alpha)$, we would have $p_c \leq \max\{p_a, p_b, p_d\}$, which in turn means that the total effect c must be significant if a, b and d are all significant. In this paper, we show via the following theorem that the rejection regions indeed enjoy such as nice geometry under a mild condition.

Theorem 1. *Suppose there are n data points in the classic mediation model. Let $\mathbf{1} = (1, \ldots, 1)^T$ be the n-dimensional column vector whose elements all equal to* 1, *and let $\mathbf{X}, \mathbf{M}, \mathbf{Y}$ be the column data vectors for variables X, M and Y respectively. Let $\mathcal{D} = (\mathbf{1}, \mathbf{X}, \mathbf{M}, \mathbf{Y})$ denote the data matrix of the regression. If* $\mathrm{rank}(\mathcal{D}) = 4$, *then the condition $\hat{a} \times \hat{b} \times \hat{d} > 0$ implies* $\mathrm{sign}(\hat{c}) = \mathrm{sign}(\hat{d})$ *and*

$$\mathcal{R}_a(\alpha) \cap \mathcal{R}_b(\alpha) \cap \mathcal{R}_d(\alpha) \subseteq \mathcal{R}_c(\alpha) \textit{ for all } \alpha \in (0,1).$$

To verify Theorem 1, we need to derive the concrete form of the involved LSEs and rejection regions. For a multivariate linear regression problem

$$Y = \beta_0 X_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \varepsilon \tag{3.1}$$

with n data points $\{(X_{i0}, X_{i1}, \ldots, X_{ip}, Y_i)\}_{i=1}^n$, we let $\hat{\beta}$ be the LSE of $\beta = (\beta_0, \ldots, \beta_p)$, and let $\mathcal{R}_j(\alpha)$ be the level-α rejection region of the F-test for testing hypotheses

$$H_0 : \beta_j = 0 \quad \text{vs} \quad H_1 : \beta_j \neq 0. \tag{3.2}$$

Let $\mathbf{Y} = (Y_1, \ldots, Y_n)^T$ and $\mathbf{X}_j = (X_{1j}, \ldots, X_{nj})^T$ be the response vector and the jth predictor vector, respectively. We write the design matrix as $\mathbf{X} = (\mathbf{X}_0, \mathbf{X}_1, \ldots, \mathbf{X}_p)$, and denote

$$\mathbf{X}[-j] = (\mathbf{X}_0, \ldots, \mathbf{X}_{j-1}, \mathbf{X}_{j+1}, \ldots, \mathbf{X}_p) \text{ for all } j \in \{0, \ldots, p\}.$$

The classic theory for linear regression (Neter, Wasserman and Kutner (1989)) tells us that

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}, \tag{3.3}$$

$$\mathcal{R}_j(\alpha) = \left\{ (\mathbf{X}, \mathbf{Y}) : \frac{||\mathbf{Y}_{\mathbf{X}} - \mathbf{Y}_{\mathbf{X}[-j]}||/1}{||\mathbf{Y} - \mathbf{Y}_{\mathbf{X}}||/(n-p-1)} > \lambda_{1,n-p-1}(\alpha) \right\}, \tag{3.4}$$

where $\mathbf{Y}_{\mathbf{X}}$ stands for the projection of vector $\mathbf{Y}$ onto the linear space, $\mathrm{span}(\mathbf{X})$, and $\lambda_{t,s}(\alpha)$ represents the αth-quantile of F-distribution with the degrees of free-

Table 1. Tests to establish mediation: models, hypotheses and rejection regions of the F-tests for each parameter.

Test	Model	Hypotheses	Rejection region of F-test
$\mathcal{T}_a$	$M = i_M + aX + \varepsilon_M$	$H_0 : a = 0, H_1 : a \neq 0$	$\mathcal{R}_a(\alpha) = \left\{ \frac{\lvert\lvert\mathbf{M_{1,X}} - \mathbf{M_1}\rvert\rvert/(2-1)}{\lvert\lvert\mathbf{M} - \mathbf{M_{1,X}}\rvert\rvert/(n-2)} > \lambda_{1,n-2}(\alpha) \right\}$
$\mathcal{T}_b$	$Y = i_Y + bM + dX + \varepsilon_Y$	$H_0 : b = 0, H_1 : b \neq 0$	$\mathcal{R}_b(\alpha) = \left\{ \frac{\lvert\lvert\mathbf{Y_{1,M,X}} - \mathbf{Y_{1,X}}\rvert\rvert/(3-2)}{\lvert\lvert\mathbf{Y} - \mathbf{Y_{1,M,X}}\rvert\rvert/(n-3)} > \lambda_{1,n-3}(\alpha) \right\}$
$\mathcal{T}_d$	$Y = i_Y + bM + dX + \varepsilon_Y$	$H_0 : d = 0, H_1 : d \neq 0$	$\mathcal{R}_d(\alpha) = \left\{ \frac{\lvert\lvert\mathbf{Y_{1,M,X}} - \mathbf{Y_{1,M}}\rvert\rvert/(3-2)}{\lvert\lvert\mathbf{Y} - \mathbf{Y_{1,M,X}}\rvert\rvert/(n-3)} > \lambda_{1,n-3}(\alpha) \right\}$
$\mathcal{T}_c$	$Y = i^*_Y + cX + \varepsilon^*_Y$	$H_0 : c = 0, H_1 : c \neq 0$	$\mathcal{R}_c(\alpha) = \left\{ \frac{\lvert\lvert\mathbf{Y_{1,X}} - \mathbf{Y_1}\rvert\rvert/(2-1)}{\lvert\lvert\mathbf{Y} - \mathbf{Y_{1,X}}\rvert\rvert/(n-2)} > \lambda_{1,n-2}(\alpha) \right\}$

dom (t, s). Applying (3.4) to the mediation model (1.1)-(1.3), we obtain the rejection regions $\mathcal{R}_a(\alpha)$, $\mathcal{R}_b(\alpha)$, $\mathcal{R}_d(\alpha)$, and $\mathcal{R}_c(\alpha)$ for the corresponding F-tests, as summarized in Table 1.

3.2. Simplifying the problem via orthogonal data transformation

Although (3.3)-(3.4) and Table 1 provide the mathematical formulation of LSEs and rejection regions of interest, it is inconvenient to verify Theorem 1 directly based on them. To further simplify the problem, we note that the statistical inference of β in terms of LSE and F-tests does not depend on the choice of the coordinate system in the data space of the regression model as stated by the lemma below:

Lemma 1. *Let* $\mathcal{D} = (\mathbf{X}_0, \mathbf{X}_1, \ldots, \mathbf{X}_p, \mathbf{Y})$ *be the data matrix of regression model* (3.1)*. For any* $n \times n$ *real orthogonal matrix* Γ *satisfying* $\Gamma'\Gamma = I_n$ *and global scale parameter* $\gamma > 0$*, define* $\tilde{\mathcal{D}} = (\tilde{\mathbf{X}}_0, \tilde{\mathbf{X}}_1, \ldots, \tilde{\mathbf{X}}_p, \tilde{\mathbf{Y}}) = \gamma\Gamma\mathcal{D}$ *be the transformed data matrix and*

$$\tilde{Y} = \beta_0 \tilde{X}_0 + \beta_1 \tilde{X}_1 + \cdots + \beta_p \tilde{X}_p + \varepsilon \tag{3.5}$$

be the transformed regression problem. Let $\tilde{\beta}$ *be the LSE of* β *and* $\tilde{\mathcal{R}}_j(\alpha)$ *be the corresponding rejection region of* F*-test for hypotheses* (3.2) *under the transformed problem* (3.5)*. We have*

$$\tilde{\beta} = \hat{\beta} \text{ and } \tilde{\mathcal{R}}_j(\alpha) = \mathcal{R}_j(\alpha) \text{ for all } j \in \{0, \ldots, p\} \text{ and } \alpha \in (0, 1).$$

Lemma 1 means that we can choose a convenient coordinate system to work with in a regression model without changing the results of statistical inference for regression coefficient. Considering that data matrix $\mathcal{D} = (\mathbf{1}, \mathbf{X}, \mathbf{M}, \mathbf{Y})$ in the classic mediation model and $rank(\mathcal{D}) = 4$, the four column vectors in $\mathcal{D}$ span a

4-dimensional subspace in $\mathbb{R}^n$. With the freedom to reset the coordinate system of $\mathbb{R}^n$ and the scale of the four data vectors, we can certainly find an orthogonal coordinate system of the data space under which the vector representation of the four original data vectors becomes $\tilde{\mathbf{1}} = (1, 0, \ldots, 0)^T$, $\tilde{\mathbf{X}} = (x_1, x_2, 0, \ldots, 0), \tilde{\mathbf{M}} = (m_1, m_2, m_3, 0, \ldots, 0), \tilde{\mathbf{Y}} = (y_1, y_2, y_3, y_4, 0, \ldots, 0)$ with $x_2 > 0, m_3 > 0$ and $y_4 > 0$. Let $\tilde{\mathcal{D}} = (\tilde{\mathbf{1}}, \tilde{\mathbf{X}}, \tilde{\mathbf{M}}, \tilde{\mathbf{Y}})$ be the data matrix under the new coordinate system. Clearly, $\tilde{\mathcal{D}}$ is an upper triangular matrix.

Because different coordinate systems can be mapped to each other via orthogonal transformations, we can also interpret $\tilde{\mathcal{D}}$ as a transformation of the original data matrix $\mathcal{D}$, i.e., there exists an orthogonal matrix Q such that $\tilde{\mathcal{D}} = \gamma Q'\mathcal{D}$, where the factor $\gamma = 1/\sqrt{n}$ rescales the vector $\mathbf{1}$ to have unit length. In theory, the configuration of orthogonal matrix Q is typically not unique, as there are often more than one coordinate systems that satisfy our conditions. In practice, however, we can always find a specific configuration of Q via the standard Gram-Schmidt process. We detail the process in Supplementary Materials. Since Lemma 1 ensures that $\mathcal{D}$ and $\tilde{\mathcal{D}}$ lead to the exactly the same LSEs and rejection regions for (a, b, d, c), and projection calculation becomes much easier for the transformed data matrix $\tilde{\mathcal{D}}$, we can derive an explicit form of LSEs and geometric shapes of rejection regions in Lemma 2.

Lemma 2. *Based on the transformed data matrix $\tilde{\mathcal{D}}$, we have:*

$$\hat{a} = \tilde{a} = \frac{m_2}{x_2}, \quad \hat{b} = \tilde{b} = \frac{y_3}{m_3}, \quad \hat{c} = \tilde{c} = \frac{y_2}{x_2}, \quad \hat{d} = \tilde{d} = \frac{(m_3 y_2 - m_2 y_3)}{x_2 m_3};$$

$$\begin{aligned}
\mathcal{R}_a(\alpha) &= \tilde{\mathcal{R}}_a(\alpha) = \{r > r_{n,\alpha}\}, \\
\mathcal{R}_b(\alpha) &= \tilde{\mathcal{R}}_b(\alpha) = \{p > p_{n,\alpha}\}, \\
\mathcal{R}_c(\alpha) &= \tilde{\mathcal{R}}_c(\alpha) = \left\{q > r_{n,\alpha}(p^2+1)^{1/2}\right\}, \\
\mathcal{R}_d(\alpha) &= \tilde{\mathcal{R}}_d(\alpha) = \begin{cases} \{|q - rp| > p_{n,\alpha}(r^2+1)^{1/2}\}, & \text{if } \hat{a}\hat{b}\hat{c} \geq 0, \\ \{|q + rp| > p_{n,\alpha}(r^2+1)^{1/2}\}, & \text{if } \hat{a}\hat{b}\hat{c} < 0; \end{cases}
\end{aligned}$$

where $r = |m_2|/m_3$, $p = |y_3|/y_4$ and $q = |y_2|/y_4$; $r_{n,\alpha} = [\lambda_{1,n-2}(\alpha)/(n-2)]^{1/2}$ and $p_{n,\alpha} = [\lambda_{1,n-3}(\alpha)/(n-3)]^{1/2}$ are constants for fixed sample size n and significant level α.

Lemma 2 indicates that each of the four rejection regions of interest corresponds to a subspace in a 3-dimensional space indexed by (p, q, r), which degenerates to a region in p-q plane $\mathcal{P}_r$ for each specific value of r. Let $\mathcal{R}_j(\alpha|r)$ be the intersection of $\mathcal{R}_j(\alpha)$ and $\mathcal{P}_r$ for all $j \in \{a, b, c, d\}$. Apparently, $\mathcal{R}_a(\alpha|r) = \mathcal{P}_r \bigcap I(r > r_{n,\alpha})$ corresponds to either an empty set or the whole p-q plane

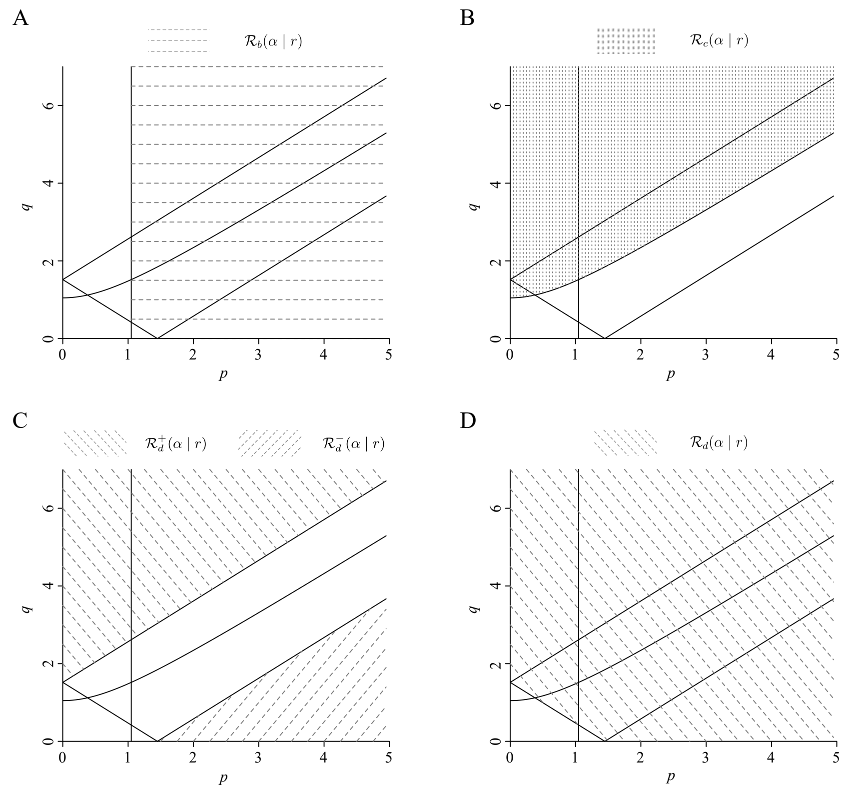

Figure 2. A graphical illustration of $\mathcal{R}_b(\alpha|r), \mathcal{R}_c(\alpha|r)$ and $\mathcal{R}_d(\alpha|r)$ in the p-q for a fixed r: (A) $\mathcal{R}_b(\alpha|r)$, (B) $\mathcal{R}_c(\alpha|r)$, (C) $\mathcal{R}_d(\alpha|r)$ when $\hat{a}\hat{b}\hat{c} \geq 0$, (D) $\mathcal{R}_d(\alpha|r)$ when $\hat{a}\hat{b}\hat{c} < 0$.

$\mathcal{P}_r$ depending on the value of r. Region $\mathcal{R}_b(\alpha|r)$ is the right half of $\mathcal{P}_r$ beyond the vertical line $p = p_{n,\alpha}$. Region $\mathcal{R}_c(\alpha|r)$ corresponds the space above the higher branch of the hyperbola with asymptotes $q = \pm\, r_{n,\alpha} p$ and vertices $(0, \pm\, r_{n,\alpha})$. The structure of region $\mathcal{R}_d(\alpha|r) = \mathcal{R}_d^+(\alpha|r) \cup \mathcal{R}_d^-(\alpha|r)$, however, is a bit complicated. When $\hat{a}\hat{b}\hat{c} \geq 0$, $\mathcal{R}_d(\alpha|r)$ contains two disconnected subregions $\mathcal{R}_d^+(\alpha|r)$ and $\mathcal{R}_d^-(\alpha|r)$, where $\mathcal{R}_d^+(\alpha|r) = \{q > p_{n,\alpha}\sqrt{r^2+1} + rp\}$ being the region above the straight line with intercept $t_{r,\alpha} = p_{n,\alpha}\sqrt{r^2+1}$ and slope $k_r = r$, and $\mathcal{R}_d^-(\alpha|r) = \{q < -p_{n,\alpha}\sqrt{r^2+1} + rp\}$ being the region below the straight line with intercept $-t_{r,\alpha}$ and slope $k_r = r$. When $\hat{a}\hat{b}\hat{c} < 0$, however, the two components of $\mathcal{R}_d(\alpha|r)$ change accordingly into the new forms below: $\mathcal{R}_d^+(\alpha|r) = \{q > p_{n,\alpha}\sqrt{r^2+1} - rp\}$ and $\mathcal{R}_d^-(\alpha|r) = \{q < -p_{n,\alpha}\sqrt{r^2+1} - rp\}$, with $\mathcal{R}_d^-(\alpha|r)$ vanishes due to the constraints that $p > 0$ and $q > 0$. Figure 2 provides a graphical demonstration for the geometry of $\mathcal{R}_b(\alpha|r)$, $\mathcal{R}_c(\alpha|r)$ and the effective components of $\mathcal{R}_d(\alpha|r)$ under different conditions, respectively.

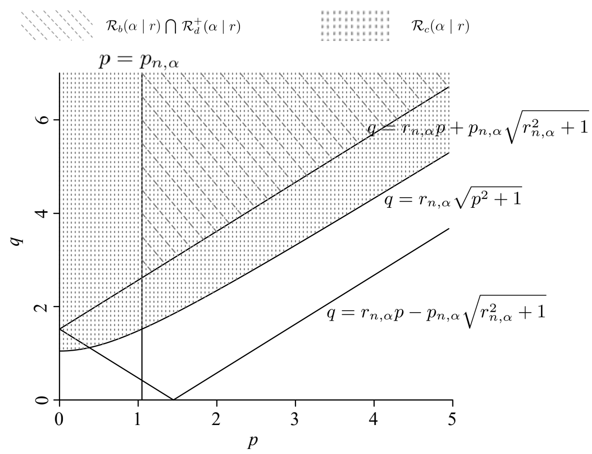

Figure 3. Geometry of $\mathcal{R}_b(\alpha|r) \bigcap \mathcal{R}_d(\alpha|r)$ and $\mathcal{R}_c(\alpha|r)$ in p-q plane for complementary mediation.

3.3. Geometric analysis for complementary mediation

To claim complementary mediation, we require $\hat{a}\hat{b}\hat{d} > 0$ as a necessary condition.

Corollary 1. *If* $\hat{a}\hat{b}\hat{d} > 0$, *we have:* (1) $\mathrm{sign}(\hat{c}) = \mathrm{sign}(\hat{d})$, *and* (2) $\mathcal{R}_d^-(\alpha) = \emptyset$, *and thus* $\mathcal{R}_d(\alpha) = \mathcal{R}_d^+(\alpha) = \{q > rp + p_{n,\alpha}(r^2+1)^{1/2}\}$.

Based on the above reasoning, for complementary mediation, the geometry of $\mathcal{R}_a(\alpha|r)$, $\mathcal{R}_b(\alpha|r)$, $\mathcal{R}_d(\alpha|r)$ and $\mathcal{R}_c(\alpha|r)$ can be demonstrated as in Figure 3. Obviously, Theorem 1 holds if and only if

$$\mathcal{R}_a(\alpha|r) \cap \mathcal{R}_b(\alpha|r) \cap \mathcal{R}_d(\alpha|r) \subseteq \mathcal{R}_c(\alpha|r) \text{ for all } \alpha \in (0,1) \text{ and } r \in (0,+\infty). \quad (3.6)$$

As (3.6) trivially holds for all $r \le r_{n,\alpha}$, we only need to consider the scenario where $r > r_{n,\alpha}$. In this case, the geometry in Figure 3 shows that a sufficient and necessary condition of (3.6) is: the boundary of $\mathcal{R}_b(\alpha|r) \cap \mathcal{R}_d(\alpha|r)$ stays away from the boundary of $\mathcal{R}_c(\alpha|r)$ for all $\alpha \in (0,1)$, which is ensured by the condition: for all $n > 3$ and $\alpha \in (0,1)$,

$$\pi_{n,\alpha} = t_{r_{n,\alpha}} - r_{n,\alpha} = p_{n,\alpha}(r_{n,\alpha}^2 + 1)^{1/2} - r_{n,\alpha} \ge 0. \quad (3.7)$$

The Lemma below guarantees that inequality (3.7) holds. Therefore, we complete the proof of Theorem 1.

Lemma 3. *For all* $n > 3$ *and* $\alpha \in (0,1)$, $p_{n,\alpha} \ge r_{n,\alpha}$.

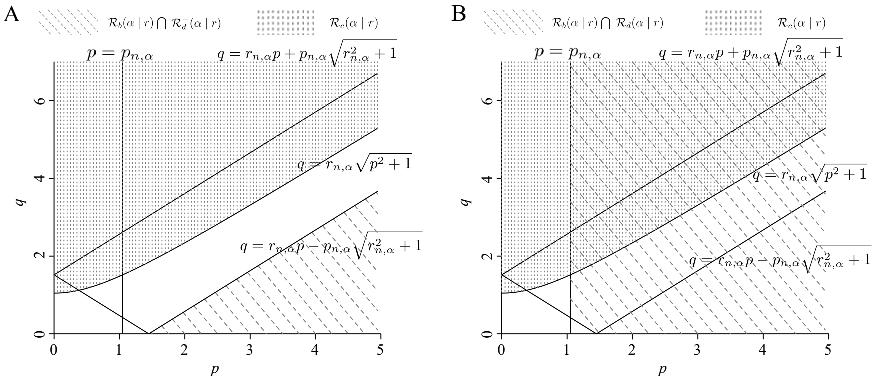

Figure 4. Geometry of $\mathcal{R}_b(\alpha|r) \bigcap \mathcal{R}_d(\alpha|r)$ and $\mathcal{R}_c(\alpha|r)$ in p-q plane for competitive mediation: (A) when $\hat{a}\hat{b}\hat{c} \geq 0$, (B) when $\hat{a}\hat{b}\hat{c} < 0$.

3.4. Impact on analysis of complementary mediation

The above results suggest that total-effect test is superfluous for establishing complementary mediation under the LSE-F framework. To establish complementary mediation, only the following are needed:

1. obtain $\hat{a}, \hat{b}, \hat{d}$ via LSE;
2. establish $\hat{a}\hat{b}\hat{d} > 0$;
3. establish $\hat{a}, \hat{b}$ and $\hat{d}$ each is statistically significant via the standard F-test for regression coefficients;

Complementary mediation is established if all of the above three are satisfied.

3.5. Extension to mediation of other types

Similar geometric analyses can be adopted to study other types of mediation. It has been generally accepted that total-effect test is unnecessary for establishing competitive mediation, because the mediated effect and the direct effect may offset each other to produce a non-significant total effect. To support the argument via geometric analysis, we only need to show that $\mathcal{R}_b(\alpha|r) \bigcap \mathcal{R}_d(\alpha|r)$ and $\mathcal{R}_c(\alpha|r)$ do not bother each other in general.

Figure 4 demonstrates the geometry of $\mathcal{R}_b(\alpha|r) \bigcap \mathcal{R}_d(\alpha|r)$ and $\mathcal{R}_c(\alpha|r)$ when $\hat{a}\hat{b}\hat{d} < 0$. Because the value of $\hat{a}\hat{b}\hat{c}$ can be positive or negative in this case (condition $\hat{a}\hat{b}\hat{d} < 0$ does not necessarily lead to a positive or negative $\hat{a}\hat{b}\hat{c}$ as in the complementary mediation), the geometry of $\mathcal{R}_d(\alpha)$ has two alternative forms depending on the sign of $\hat{a}\hat{b}\hat{c}$ based on Lemma 2 and needs to be discussed sepa-

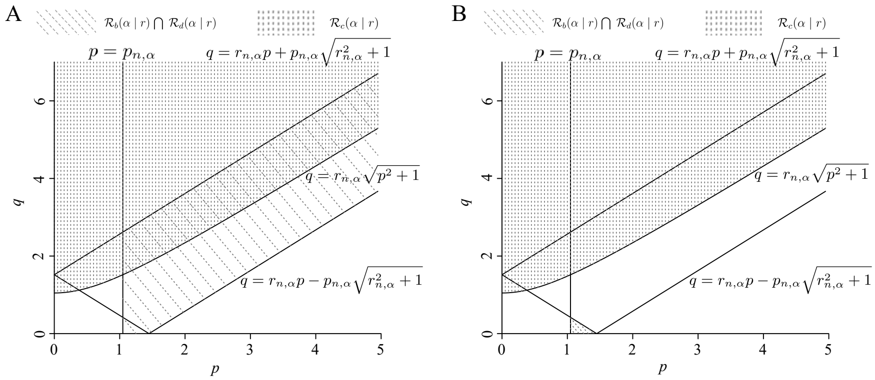

Figure 5. Geometry of $\mathcal{R}_b(\alpha|r) \bigcap \mathcal{R}_d^c(\alpha|r)$ and $\mathcal{R}_c(\alpha|r)$ in p-q plane for indirect-only mediation: (A) when $\hat{a}\hat{b}\hat{c} \geq 0$, (B) when $\hat{a}\hat{b}\hat{c} < 0$.

rately. Figure 4 (A) and (B) correspond to each of the two scenarios, respectively. From these figures, we can see that $\mathcal{R}_a(\alpha) \cap \mathcal{R}_b(\alpha) \cap \mathcal{R}_d(\alpha)$ and $\mathcal{R}_c(\alpha)$ can either completely separate from each other (when $\hat{a}\hat{b}\hat{c} > 0$) or share a common subregion (when $\hat{a}\hat{b}\hat{c} < 0$), confirming that the total effect test is indeed irrelevant to establishing a competitive mediation.

Similarly, we expect an unconstrained relationship between $\mathcal{R}_a(\alpha) \cap \mathcal{R}_b(\alpha) \cap \mathcal{R}_d^c(\alpha)$ and $\mathcal{R}_c(\alpha)$ for indirect-only mediation, as it has been suspected that total-effect may not be significant here. Figure 5 demonstrates the geometry of $\mathcal{R}_b(\alpha|r) \bigcap \mathcal{R}_d^c(\alpha|r)$ and $\mathcal{R}_c(\alpha|r)$ with no further constraints on LSEs. The figures show that there exist cases where $\mathcal{R}_b(\alpha|r) \bigcap \mathcal{R}_d^c(\alpha|r)$ and $\mathcal{R}_c(\alpha|r)$ intersect but do not contain each other, as in Figure 5 (A), or completely separate from each other, as in Figure 5 (B), confirming an unconstrained relationship between $\mathcal{R}_a(\alpha) \cap \mathcal{R}_b(\alpha) \cap \mathcal{R}_d^c(\alpha)$ and $\mathcal{R}_c(\alpha)$ in general. These results suggest that the geometric analysis proposed in this paper could serve as a general tool for studying mediation of various types.

4. Simulation Studies

4.1. Numerical validation of Theorem 1

The theoretical result above the total-effect test under the LSE-F framework, as stated in Theorem 1, can be validated numerically by simulation. For this purpose, we generated simulated data from the mediation model (1.1) and (1.2) as follows:

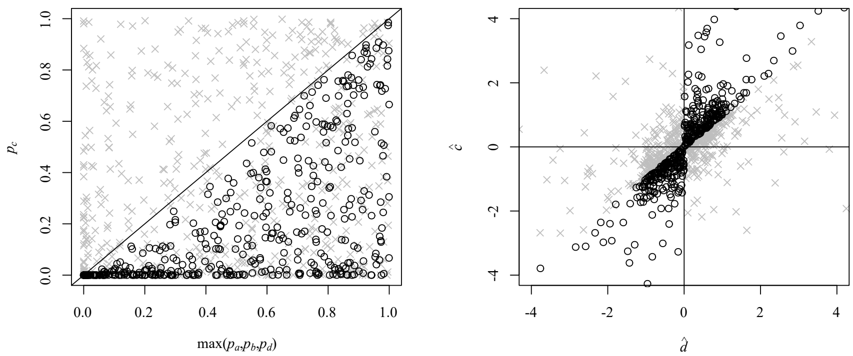

Figure 6. Numerical validation of Theorem 1: black circles represent datasets with $\hat{a}\hat{b}\hat{d} > 0$, and grey crossings represent datasets with $\hat{a}\hat{b}\hat{d} \leq 0$.

$$n \sim \text{Unif}(\{4, \ldots, 100\}),\ (i_M, i_Y, a, b, d) \sim \text{Unif}[-1, 1]^5,$$
$$X \sim N(0, 1),\ \sigma_M^2 \text{ and } \sigma_Y^2 \sim \text{Inv-Gamma}(1, 1).$$

A total of 1,000 independent datasets of different sample sizes were simulated for numerical validation.

For each simulated dataset, we calculated the LSEs $(\hat{a}, \hat{b}, \hat{d}, \hat{c})$ and p-values (p_a, p_b, p_d, p_c) under the LSE-F framework. If our theory holds, we would expect to see $p_c \leq \max\{p_a, p_b, p_d\}$ and $\hat{d}\hat{c} > 0$ for all runs in which $\hat{a}\hat{b}\hat{d} > 0$. Figure 6 checks the above expectations in a graphical manner. Figure 6 (A) checks the p-value condition by demonstrating each simulated dataset with one point in a 2-dimensional space with the X-axis representing $\max\{p_a, p_b, p_d\}$, the Y-axis standing for p_c, and the shape highlighting the type of points: black circles for datasets satisfying $\hat{a}\hat{b}\hat{d} > 0$, and grey crossings for all the other datasets; Figure 6 (B) checks the estimator sign condition in a similar way with the X-axis and Y-axis representing $\hat{d}$ and $\hat{c}$, respectively. We can see from these figures that, although the grey crossing points spread all over the figures, all black circle points are located under the diagonal line in Figure 6 (A), and within the up-right and down-left quadrants in Figure 6 (B). These results are consistent with our expectation, and thus support our theory.

4.2. Exploratory analysis for other frameworks

To explore whether a similar result holds for other frameworks for establishing complementary mediation, we also implemented a similar numerical analysis for

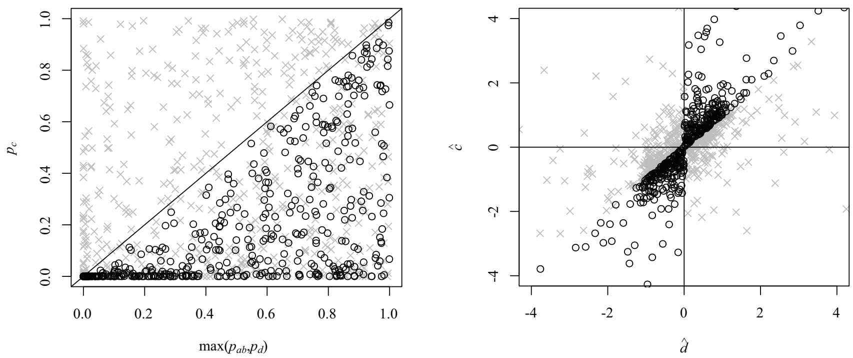

Figure 7. Numerical exploration for LSE-Sobel framework: black circles represent datasets with $\hat{a}\hat{b}\hat{d} > 0$, and grey crossings represent datasets with $\hat{a}\hat{b}\hat{d} \leq 0$.

LSE-Sobel framework and LAD-Z framework with the same group of simulated datasets. For the LSE-Sobel framework, we calculated the LSEs $(\hat{a}, \hat{b}, \hat{d}, \hat{c})$ for each simulated dataset, and p-values of the corresponding tests, including p_{ab}, the p-value of the Sobel test for $a \times b$, p_d, the p-value of the F-test for d, and p_c, the p-value of the F-test for c. If a similar result holds for the LSE-Sobel framework, we would expect to see that $p_c \leq \max\{p_{ab}, p_d\}$ and $\hat{d}\hat{c} > 0$ for all runs in which $\hat{a}\hat{b}\hat{d} > 0$.

Very similar to Figure 6, Figure 7 provides a graphical demonstration of the results under the LSE-Sobel framework. Clearly, all black circle points are located under the diagonal line in Figure 7 (A), suggesting that LSE-Sobel framework may share a similar property as LSE-F framework, i.e., $p_c \leq \max\{p_{ab}, p_d\}$. Furthermore, considering that Figure 7 (B) is exactly the same as Figure 6 (B), as LSE-Sobel framework and LSE-F framework are identical in parameter estimation, we tend to believe that the total-effect test is superfluous under the LSE-Sobel framework as well. In fact, the theoretical result below provides us confidence about this conjecture when sample size n is large enough.

Theorem 2. *Let p_{ab} be the p-value of the Sobel test, and let p_a and p_b be p-values of F-tests for a and b, respectively. Then, for all $\varepsilon > 0$, there exists $N > 0$ such that as long as the sample size $n > N$, we have $p_{ab} \geq \max\{p_a, p_b\} - \varepsilon$.*

Theorem 2 leads to the following corollary immediately.

Corollary 2. *If $\hat{a}\hat{b}\hat{d} > 0$, then*

$$\operatorname{sign}(\hat{c}) = \operatorname{sign}(\hat{d}) \text{ and } \lim_{n\to\infty} P\big(p_c \leq \max\{p_{ab}, p_d\}\big) = 1.$$

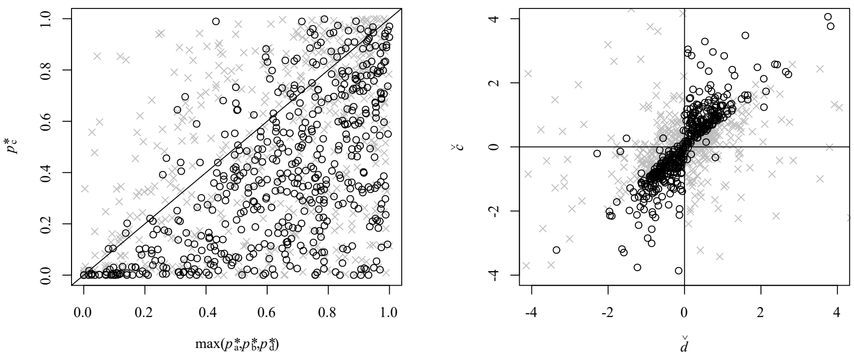

Figure 8. Numerical exploration for LAD-Z framework: black circles represent datasets with $\check{a}\check{b}\check{d} > 0$, and grey crossings represent datasets with $\check{a}\check{b}\check{d} \leq 0$.

For LAD-Z framework, however, the similar property does not hold. Let $(\check{a}, \check{b}, \check{d}, \check{c})$ be the LADs of model parameters (a, b, d, c), and let $(p_a^*, p_b^*, p_d^*, p_c^*)$ be the p-values of the corresponding Z-tests. Figure 8 shows the scatter plots of $(\max\{p_a^*, p_b^*, p_d^*\}, p_c^*)$ and $(\check{d}, \check{c})$ based on the 1,000 simulation datasets in a similar fashion as Figure 6 and Figure 7. We also replaced the Gaussian errors in (1.1, 1.2) by Laplace errors and the figures have the same pattern. Unfortunately, we find that some black circles (less than 10%), which represent the datasets with $\check{a}\check{b}\check{d} > 0$, spread over the diagonal line, suggesting that there is no easy answer to the role of total-effect test under the LAD-Z framework.

5. Real Data Applications

To illustrate our main thesis with real data, we reanalyzed responses to a 1987 opinion survey, which asked 870 randomly selected Beijing residents about their attitudes toward the economic reform under debate (Zhao et al. (1994)). The dataset is of historical significance. The survey is one of the first in China mainland based on probability sampling. It provides a rare view of the public opinion at the very beginning of the reform, which in the following decades transformed one of the poorest economies into the second largest in the world (Zhao et al. (1994); Zhao and Shen (1995); Chen et al. (2008)). Hereinafter, we refer to this dataset as the *Opinion-1987 Data*.

Our reanalysis focuses on how the media affected Beijingers' understanding of the reasons for the reform, and how the understanding in turn affected Beijingers' support for the reform, i.e., Use-Media→ Understand-Reason→ Support-Reform.

Table 2. Descriptive statistics of variables in the Opinion-1987 Data: the sample size N, the original scale as data were collected and the 0-1 percentage scale after the variables have been linearly transformed to the interval $[0, 1]$.

		N	Original Scale				0-1 Percentage Scale			
			Min	Max	Mean	Sd	Min	Max	Mean	Sd
Y	Support_Reform	847	1	5	4.28	0.79	0	1	0.821	0.196
M	Understand_Reason	846	1	5	3.84	0.88	0	1	0.709	0.220
X	Read_Paper (days/10 days)	838	0	10	5.43	3.59	0	1	0.543	0.359
	Listen_to_Radio (days/10 days)	842	0	10	5.45	3.64	0	1	0.545	0.364
	Watch_TV (days/10 days)	844	0	10	6.25	3.40	0	1	0.625	0.340
	Use_Media (days/10 days)	844	0	10	5.71	2.69	0	1	0.571	0.269

The data and the variables were described in detail by Zhao et al. (1994). Below we highlight some information for this reanalysis.

5.1. Variables in the data

Dependent variable: support for reform (Support-Reform). This is a weighted average of the responses to three questions measuring respondents' attitude toward the government's economic policy, originally on 5-point Likert scales. For this reanalysis, the composite variable was linearly transformed to a 0-1 percentage scale where 1 represents the strongest support and 0 represents the strongest opposition (Zhao et al. (1994); Zhao and Zhang (2014)).

Mediating variable: understanding reasons (Understand-Reason). This is a weighted average of the responses to seven questions measuring respondents' acceptance of the reasons in support of the reform, originally also on 5-point Likert scales. Again, the composite variable was linearly transformed to a 0-1 percentage scale where 1 represents the strongest acceptance and 0 represents the strongest rejection (Zhao et al. (1994); Zhao and Zhang (2014)).

Independent variables: media exposure (Read-Paper, Listen-to-Radio, Watch-TV and Use-Media). Three variables measured how often the respondents read newspaper, listened to radio or watched television. A fourth variable, Use-Media, was created by taking the average of the three. Each of the four was transformed to a 0-1 percentage scale where 1 represents exposure every day, and 0 represents no exposure at all.

Univariate descriptions of all variables are in Table 2. The original Opinion-1987 Data also contains seven control variables, including Age and Education. We omitted the control variables to simplify the analysis.

Table 3. Mediation analysis results of the Opinion-1987 Data. Columns 2-6: the independent variable name; LSEs of the parameters; whether the condition $\hat{a}\hat{b}\hat{d} > 0$ holds; p-values of testing each parameter and mediation types of each model. The mediation type is determined by fixing significance level $\alpha = 0.05$. Note that, as all variables are on 0-1 percentage scales, all regression coefficients, namely $\hat{a}, \hat{b}, \hat{d}$, and $\hat{c}$, become percentage coefficients (bp) (Zhao and Zhang (2014)).

Model	X	Estimates				$I(\hat{a}\hat{b}\hat{d} > 0)$	p-values				Mediation Type
		$\hat{a}$	$\hat{b}$	$\hat{d}$	$\hat{c}$		p_a	p_b	p_d	p_c	
$\mathcal{M}_1$	Read_Paper	0.229	0.239	0.048	0.102	Yes	<2e-16	1.53e-13	1.46e-2	6.54e-8	Complementary
$\mathcal{M}_2$	Listen_to_Radio	0.155	0.242	0.059	0.096	Yes	4.65e-14	3.40e-15	1.23e-3	1.99e-7	Complementary
$\mathcal{M}_3$	Watch_TV	0.025	0.265	0.034	0.041	Yes	0.271	<2e-16	7.44e-2	4.19e-2	Non-mediation
$\mathcal{M}_4$	Use_Media	0.242	0.238	0.080	0.137	Yes	<2e-16	1.88e-14	1.37e-3	3.94e-8	Complementary

5.2. Mediation analysis under various models

By alternating the four independent variables while retaining the same dependent and mediating variables, we constructed four models of potential mediation. Table 3 summarizes the four models and results of corresponding mediation analysis. From the table, we can see that complementary mediation shows up in three of the four models, i.e., models $\mathcal{M}_1, \mathcal{M}_2$ and $\mathcal{M}_4$, while no mediation effect is found in model $\mathcal{M}_3$, probably due to the relatively low television penetration in China at the time. Theorem 1 holds for all these real datasets, and the total-effect test is indeed superfluous, as predicted by Theorem 1.

6. Conclusion and Discussion

This article provides an explicit proof that the total effect always bears the same sign as the direct effect and is statistically significant whenever the mediated effect and the direct effect point to the same direction and are both significant, as long as LSE and F-tests are used to establish mediation, therefore is superfluous and unnecessary for establishing mediation of this type in the classic mediation model. We also show by numerical study and theoretical analysis that the similar result also holds for the LSE-Sobel framework when sample size is large enough. Considering that total-effect test can erroneously reject competitive mediation and indirect-only mediation, our finding supports the growing agreement that the total-effect test is unnecessary for establishing any type of mediation.

The discussions in this study are limited to the classic mediation model, where X and M influence Y and each other linearly. For the more general cases where X and M influence Y in a non-linear way with interactions, it becomes conceptually tricky and technically more challenging to define, estimate, and test for mediation effects. See Robins and Greenland (1992), Pearl (2001), Frangakis and Rubin (2002), Lindquist (2012) for different extensions of the classic direct or indirect effect in a general setting, and Pearl (2012), Daniels et al. (2012) for estimation methods. More efforts are needed to study the role of total-effect test in the more general settings.

Supplementary Material

Supplementary materials available online include the details for constructing the transformed data matrix $\tilde{\mathcal{D}}$ and a detailed proof for Lemma 2.

Acknowledgements

This research is supported in part by National Natural Science Foundation of China (Grant 11771242, DK PI), Beijing Academy of Artificial Intelligence (Supporting Grant BAAI2019ZD0103, DK PI), Higher Education Fund of the Government of Macao SAR (ZXS PI), and grants of University of Macau, including FSS One-off Research Assistance Fund, CPG2018-00035-FSS, SRG2018-00143-FSS (ZXS PI), and CGBAS, UMDF-00712020 (HWX PI).

Appendix

A. Proof of Lemma 1

Apparently,

$$\tilde{\beta} = (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{Y}} = [(\gamma\Gamma\mathbf{X})'(\gamma\Gamma\mathbf{X})]^{-1}(\gamma\Gamma\mathbf{X})'(\gamma\Gamma\mathbf{Y}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'Y = \hat{\beta};$$

and, $\mathcal{R}_j(\alpha) = \tilde{\mathcal{R}}_j(\alpha)$ for all $j \in \{0, \ldots, p\}$ and $\alpha \in (0, 1)$ as the F-statistics is invariant under the transformation, i.e.,

$$\tilde{F}_j = \frac{||\tilde{\mathbf{Y}}_{\tilde{\mathbf{X}}} - \tilde{\mathbf{Y}}_{\tilde{\mathbf{X}}[-j]}||/1}{||\tilde{\mathbf{Y}} - \tilde{\mathbf{Y}}_{\tilde{\mathbf{X}}}||/(n-p-1)} = \frac{||\mathbf{Y}_{\mathbf{X}} - \mathbf{Y}_{\mathbf{X}[-j]}||/1}{||\mathbf{Y} - \mathbf{Y}_{\mathbf{X}}||/(n-p-1)} = F_j, \ j \in \{0, \ldots, p\}.$$

B. Proof of Lemma 2

Based on the transformed data matrix $\tilde{\mathcal{D}}$, it should be easy to see that

$$\begin{aligned}
\tilde{\mathbf{M}}_{\tilde{\mathbf{1}}} &= (m_1, 0, \ldots, 0), \ \tilde{\mathbf{M}}_{\tilde{\mathbf{1}},\tilde{\mathbf{X}}} = (m_1, m_2, 0, \ldots, 0), \\
\tilde{\mathbf{Y}}_{\tilde{\mathbf{1}}} &= (y_1, 0, \ldots, 0), \ \tilde{\mathbf{Y}}_{\tilde{\mathbf{1}},\tilde{\mathbf{X}}} = (y_1, y_2, 0, \ldots, 0), \ \tilde{\mathbf{Y}}_{\tilde{\mathbf{1}},\tilde{\mathbf{M}},\tilde{\mathbf{X}}} = (y_1, y_2, y_3, 0, \ldots, 0), \\
\tilde{\mathbf{Y}}_{\tilde{\mathbf{1}},\tilde{\mathbf{M}}} &= \left(y_1, \frac{m_2 y_2 + m_3 y_3}{m_2^2 + m_3^2} \times m_2, \frac{m_2 y_2 + m_3 y_3}{m_2^2 + m_3^2} \times m_3, 0, \ldots, 0\right).
\end{aligned}$$

Applying (3.3) and (3.4) to the transformed data, we can get the results. The detailed calculation can be found in Supplementary Material.

C. Proof of Corollary 1

Lemma 2 implies that:

$$\hat{a}\hat{b}\hat{d} = \frac{m_2y_3(m_3y_2 - m_2y_3)}{x_2^2 m_3^2},$$
$$\hat{a}\hat{b}\hat{c} = \frac{m_2y_2y_3}{x_2^2 m_3}.$$

Considering that $x_2 > 0$ and $m_3 > 0$, we have:

$$\begin{aligned}\hat{a}\hat{b}\hat{d} > 0 \iff & m_2m_3y_2y_3 > m_2^2y_3^2 \\ \implies & \begin{cases} m_2y_2y_3 > 0 & \iff \hat{a}\hat{b}\hat{c} > 0; \\ |m_3y_2| > |m_2y_3| & \iff q > rp. \end{cases}\end{aligned}$$

Note that the condition $\hat{a}\hat{b}\hat{c} > 0$ implies $\mathcal{R}_d(\alpha) = \{|q - rp| > p_{n,\alpha}(r^2 + 1)^{1/2}\}$. Furthermore, $q > rp$ implies $\mathcal{R}_d^-(\alpha) = \emptyset$ and thus $\mathcal{R}_d(\alpha) = \mathcal{R}_d^+(\alpha) = \{q > rp + p_{n,\alpha}(r^2 + 1)^{1/2}\}$.

D. Proof of Lemma 3

Let W_n follows F-distribution with the degree of freedom (1,n) and $Z_0, Z_1, \ldots, Z_m$ be a series of independent standard normal random variables, based on the definition of $\lambda_{1,n-2}(\alpha)$ and $\lambda_{1,n-3}(\alpha)$, we have: for all $\lambda > 0$,

$$P\left(\frac{W_{n-2}}{n-2} \le \frac{\lambda_{1,n-2}(\alpha)}{n-2}\right) = 1 - \alpha = P\left(\frac{W_{n-3}}{n-3} \le \frac{\lambda_{1,n-3}(\alpha)}{n-3}\right),$$
$$\begin{aligned}P\left(\frac{W_{n-3}}{n-3} \le \lambda\right) &= P\left(\frac{Z_0^2}{\sum_{i=1}^{n-3} Z_i^2} \le \lambda\right) \\ &\le P\left(\frac{Z_0^2}{\sum_{i=1}^{n-2} Z_i^2} \le \lambda\right) = P\left(\frac{W_{n-2}}{n-2} \le \lambda\right).\end{aligned}$$

As a consequence, we have

$$P\left(\frac{W_{n-2}}{n-2} \le \frac{\lambda_{1,n-2}(\alpha)}{n-2}\right) = P\left(\frac{W_{n-3}}{n-3} \le \frac{\lambda_{1,n-3}(\alpha)}{n-3}\right) \le P\left(\frac{W_{n-2}}{n-2} \le \frac{\lambda_{1,n-3}(\alpha)}{n-3}\right),$$

and thus $r_{n,\alpha} = \left(\frac{\lambda_{1,n-2}(\alpha)}{n-2}\right)^{1/2} \le \left(\frac{\lambda_{1,n-3}(\alpha)}{n-3}\right)^{1/2} = p_{n,\alpha}$.

E. Proof of Theorem 2

Let $T_a = \hat{a}^2/\mathrm{Var}(\hat{a})$ and $T_b = \hat{b}^2/\mathrm{Var}(\hat{b})$ be the test statistics of F-tests for a and b, respectively. Then the Sobel test statistic is $S^2 = 1/(1/T_a + 1/T_b)$. Let χ_1^2 be a random variable of Chi-squared distribution with degree of freedom 1 and $F_{1,n}$ be that of F-distribution with degree of freedom $(1, n)$. By definition, the p-value of Sobel test is $p_{ab} = Pr(\chi_1^2 > S^2)$ and those of F-tests are $p_a = Pr(F_{1,n} > T_a)$ and $p_b = Pr(F_{1,n} > T_b)$.

To build the relationship between p_{ab} and $\{p_a, p_b\}$, we define $\tilde{p}_a = Pr(\chi_1^2 > T_a)$ and $\tilde{p}_b = Pr(\chi_1^2 > T_b)$. Since $S^2 \le \min\{T_a, T_b\}$, we have:

$$p_{ab} = Pr(\chi_1^2 > S^2) \ge Pr(\chi_1^2 > \min\{T_a, T_b\}) = \max\{\tilde{p}_a, \tilde{p}_b\}.$$

Furthermore, $F_{1,n}$ converges to χ_1^2 in distribution, implying that as $n \to \infty$, we have uniformly convergence for all T_i that

$$p_i = Pr(F_{1,n} > T_i) \to Pr(\chi_1^2 > T_i) = \tilde{p}_i, \text{ for } i = a, b,$$

Therefore, for all $\varepsilon > 0$, there exists $N > 0$ such that for all $n > N$,

$$p_{ab} \ge \max\{\tilde{p}_a, \tilde{p}_b\} \ge \max\{p_a, p_b\} - \varepsilon.$$

F. Proof of Corollary 2

Theorem 2 shows: for all $\varepsilon > 0$, there exists $N > 0$ such that for all $n > N$, we have

$$\max\{p_a, p_b, p_d\} \le \max\{p_{ab} + \varepsilon, p_d\} \le \max\{p_{ab}, p_d\} + \varepsilon.$$

Given $\mathrm{sign}(\hat{c}) = \mathrm{sign}(\hat{d})$ and $p_c \le \max\{p_a, p_b, p_d\}$ based on Theorem 1, it is clear that the corollary holds.

References

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

Chen, Q., He, Y., Zhao, X. and Griffith, D. (2008). Sources of product information for Chinese rural consumers. *International Journal of Advertising* **27**, 67–97.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. 2nd Edition. Lawrence Erlbaum Associates, Mahwah, NJ.

Collins, L. M., Graham, J. J. and Flaherty, B. P. (1998). An alternative framework for defining mediation. *Multivariate Behavioral Research* **33**, 295–312.

Conger, A. J. (1974). A revised definition for suppressor variables: A guide to their identification and interpretation. *Educational and Psychological Measurement* **34**, 35–46.

Daniels, M. J., Roy, J. A., Kim, C., Hogan, J. W. and Perri, M. G. (2012). Bayesian inference for the causal effect of mediation. *Biometrics* **68**, 1028–1036.

Davis, J. A. (1985). *Quantitative Applications in the Social Sciences: The Logic of Causal Order.* SAGE Publications, Thousand Oaks, CA.

Frangakis, C. E. and Rubin, D. B. (2002). Principal stratification in causal inference. *Biometrics* **58**, 21–29.

Hamilton, D. (1987). Sometimes $r^2 > r^2_{yx_1} + r^2_{yx_2}$: Correlated variables are not always redundant. *The American Statistician* **41**, 129–132.

Hayes, A. F. (2009). Beyond baron and kenny: Statistical mediation analysis in the new millennium. *Communication Monographs* **76**, 408–420.

Judd, C. M. and Kenny, D. A. (1981). Process analysis: Estimating mediation in treatment evaluations. *Evaluation Review* **5**, 602–619.

Kenny, D. A., Kashy, D. A. and Bolger, N. (1998). Data analysis in social psychology. *The Handbook of Social Psychology* **1**, 233–265.

Li, Y., Baldassi, M., Johnson, E. J. and Weber, E. U. (2013). Complementary cognitive capabilities, economic decision making, and aging. *Psychology and Aging* **28**, 595–613.

Lindquist, M. A. (2012). Functional causal mediation analysis with an application to brain connectivity. *Journal of the American Statistical Association* **107**, 1297–1309.

Lord, F. M. and Novick, M. R. (2008). *Statistical Theories of Mental Test Scores.* Information Age Publishing, Charlotte, NC.

MacKinnon, D. P. and Fairchild, A. J. (2009). Current directions in mediation analysis. *Current Directions in Psychological Science* **18**, 16–20.

MacKinnon, D. P., Krull, J. L. and Lockwood, C. M. (2000). Equivalence of the mediation, confounding and suppression effect. *Prevention Science* **1**, 173–181.

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

MacKinnon, D. P., Warsi, G. and Dwyer, J. H. (1995). A simulation study of mediated effect measures. *Multivariate Behavioral Research* **30**, 41–62.

Mathieu, J. E. and Taylor, S. R. (2006). Clarifying conditions and decision points for mediational type inferences in organizational behavior. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior* **27**, 1031–1056.

McFatter, R. M. (1979). The use of structural equation models in interpreting regression equations including suppressor and enhancer variables. *Applied Psychological Measurement* **3**, 123–135.

Neter, J., Wasserman, W. and Kutner, M. H. (1989). *Applied Linear Regression Models.* Irwin Homewood, IL.

Pearl, J. (2001). Direct and indirect effects. In *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, 411–420. Morgan Kaufmann, San Francisco.

Pearl, J. (2012). The causal mediation formula-A guide to the assessment of pathways and mechanisms. *Prevention Science* **13**, 426–436. Morgan Kaufmann, San Francisco.

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

Robins, J. M. and Greenland, S. (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology* **3**, 143–155.

Rose, B. M., Holmbeck, G. N., Coakley, R. M. and Franks, E. A. (2004). Mediator and moderator effects in developmental and behavioral pediatric research. *Journal of Developmental & Behavioral Pediatrics* **25**, 58–67.

Rose, J. S., Chassin, L., Presson, C. C. and Sherman, S. J. (2000). Contrasts in multiple mediator models. In *Multivariate Applications in Substance Use Research*, 155–174. Psychology Press.

Rucker, D. D., Preacher, K. J., Tormala, Z. L. and Petty, R. E. (2011). Mediation analysis in social psychology: Current practices and new recommendations. *Social and Personality Psychology Compass* **5**, 359–371.

Shrout, P. E. and Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods* **7**, 422–445.

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological Methodology* **13**, 290–312.

Tzelgov, J. and Henik, A. (1991). Suppression situations in psychological research: Definitions, implications, and applications. *Psychological Bulletin* **109**, 524–536.

Velicer, W. F. (1978). Suppressor variables and the semipartial correlation coefficient. *Educational and Psychological Measurement* **38**, 953–958.

Wen, Z., Chang, L., Hau, K. T. and Liu, H. (2004). Testing and application of the mediating effects. *Acta Psychologica Sinica* **36**, 614–620.

Wen, Z. and Ye, B. (2014). Analyses of mediating effects: The development of methods and models. *Advances in Psychological Science* **22**, 731–745.

Zhao, X. (1997). Clutter and serial order redefined and retested. *Journal of Advertising Research* **37**, 57–74.

Zhao, X., Chen, Q. and Tong, B. (2011). Does c'test help, anytime? On communication fallacy of "effect to mediate". In *Annual Meeting of the Association for Education in Journalism and Mass Communication.* Renaissance Grand & Suites Hotel, St. Louis.

Zhao, X., Lynch Jr., J. G. and Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *Journal of Consumer Research* **37**, 197–206.

Zhao, X. and Shen, F. (1995). Audience reaction to commercial advertising in china in the 1980s. *International Journal of Advertising* **14**, 374–390.

Zhao, X. and Zhang, X. (2014). Emerging methodological issues in quantitative communication research. *New Trends in Communication Studies, II*, 953–978.

Zhao, X., Zhu, J. H., Li, H. and Bleske, G. L. (1994). Media effects under a monopoly: The case of Beijing in economic reform. *International Journal of Public Opinion Research* **6**, 95–117.

Yingkai Jiang

Yau Mathematical Sciences Center & Department of Mathematical Sciences, Tsinghua University, Haidian, Beijing, China.

E-mail: ykjiang15@gmail.com

Xinshu Zhao

Department of Communication, University of Macau, Macau, China.

School of Media and Journalism, University of North Carolina, Chapel Hill, North Carolina, USA.

E-mail: xszhao@um.edu.mo

Lixing Zhu

Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong.

E-mail: lzhu@hkbu.edu.hk

Jun S. Liu

Department of Statistics, Harvard University, Cambridge, Massachusetts, USA.

E-mail: jliu@stat.harvard.edu

Ke Deng

Center for Statistical Science & Department of Industry Engineering, Tsinghua University, Haidian, Beijing, China.

E-mail: kdeng@tsinghua.edu.cn

(Received April 2019; accepted February 2020)