Mediating Analysis with Mismeasured Variables via SIMEX Method

Sahika Gokmen[1] Rukiye Dagalp[2] and Senay Acikgoz[3]

Abstract

The casual relations among the variables of interest may be directed or transmitted by the other variable or variables. The standard mediation analysis including a series of linear regressions is used to estimate these casual direct/indirect relationships among the variables. It is well documented that the ordinary least square (OLS) estimators are biased and inconsistent under the presence of the measurement error in regressors. The simulation extrapolation (SIMEX) method is the most common method for the measurement error models. The goal of this paper is to show the success of SIMEX after the theoretical inferences about what happens to the OLS estimators with the error-prone variables. The theoretical inferences and the simulation results show that the existence of measurement error in the variables leads to biased parameter estimates. Therefore, this increases the probability of wrong decisions for testing the mediating effect via Sobel's test which is the most common test. Although the SIMEX method used to correct the biasedness of estimates, the standard errors of these estimates are obtained greater than as usual. Finally, the results suggest that the SIMEX method can be revised for the Sobel's test statistics to reveal the real mediating affect.

Keywords: Mediating analysis, Sobel's *z*-score, measurement errors, ordinary least squares, SIMEX.

JEL Classifications: C10, C12, C15, C30

[1] Econometrics, Ankara Haci Bayram Veli University, Ankara, Turkey. sahika.gokmen@hbv.edu.tr

[2] **Corresponding Author** Statistics, Ankara University, Ankara, Turkey. rdagalp@ankara.edu.tr

[3] Econometrics, Ankara Haci Bayram Veli University, Ankara, Turkey. senay.acikgoz@hbv.edu.tr

1. Introduction

Measuring is an important issue especially in social sciences that have complicated direct or indirect relationships among the variables of interest. The methods using to understand such relations among the variables in the real world are mostly affected by the existence of observations that might be obtained with the error which is called measurement error or error in variable. The ordinary least squares (OLS) estimators are among them and it is well documented that those are biased and inconsistent under the presence of measurement errors in the regressors of the models.

A relationship between two variables might be transmitted or mediated by another third variable which is called as mediator. There are different procedures using to estimates such relations and the main version for it has been defined based on estimating a series of regressions (Jude and Kenny, 1981, Barron and Kenny, 1986). This standard version of mediating analysis requires to satisfy the following assumptions in addition to the assumptions of linear regression modelling: (i) casual ordering of the explanatory (X) and response (Y) variables and mediator (M) is correct and (ii) there is no interaction effect between X and M. It is generally not enough to estimate these egressions. A test statistic defined on the several parameters of the system is used to test whether the relation from X to Y is mediated by the variable M. The Sobel's (1982) test statistics is the most popular one that is distributed approximately as normal.

As it is well known that not only in mediating analysis but also in many types of regression analysis, no errors in variables is the most ignored assumption in applications. In this recent paper, we directly focus on what happens to the OLS estimators when the mediator is mismeasured in such analysis by using simulation studies. Under the assumption of all variables in the single-mediator system are continuous with the usual assumptions of the regression analysis and mediating analysis, the simulations are repeated under small, medium and large correlations between X and M with 50, 100 and 500 sample sizes. In this recent paper, we also do simulation analysis for the Sobel's z-score for different error in variables' cases in order to see how the Sobel test statistics is affected by the measurement errors in the mediator and the explanatory variable.

Even though the number of papers using mediating effect analysis is huge, papers focusing on the measurement errors in the mediator variables are very less (Fritz et al., 2016; le Cessie et al., 2012; VanderWeele et al., 2012 and Wargas, 2015). Our paper related to our previous study has three main contributions. The first one is that measurement error might suppress the mediating effect existing or it might appear to have a mediating effect that does not actually exist. Secondly, we show that changes in the variance of the measurement error affect the estimation results under different circumstances such as a lower correlation between M and X and smaller sample size. It should be noted that disregarding the existing measurement error will lead to biased results and it requires adjusting the results by using different methods. Related to this we show that whether the simulation extrapolation (SIMEX) method, that is the most common method for the measurement error models, has benefits when there is mismeasured mediator in the analysis. The results obtained with simulation data show that although the SIMEX method produces unbiased estimates, the standard errors of these estimates are obtained greater than as usual. This evidence also implies that Sobel's test statistics will be smaller that cannot reject the null hypothesis that there is no mediating effect.

The paper has 5 sections. The second chapter summarizes the mediating analysis with a single mediator variable along with the summarized theoretical inferences about the OLS estimators under different errors in variables cases. The third section describes the SIMEX method and the fourth section gives the both simulation and SIMEX method results. The paper ends with the conclusion section.

2. The Single Mediator Model and the OLS Estimators under Measurement Error

Suppose that variables of interest are *X* and *Y*, and *Y* is determined by the variable *X*. This functional relationship including error term (*e*) that is a stochastic process indicates a casual relation form *X* to *Y* after controlling the other factors having impacts on *Y* if the covariance between *X* and *e* is zero. The marginal effect of variable *X* on *Y* may be direct or indirect. The indirect effect of *X* on *Y* may emerge with a variable that transmits the relation from *X* to *Y*. This third variable is called as a mediator (*M*) that is one of the variables of a casual chain from *X* to *Y*.

The standard mediation analysis is rooted in Baron and Kenny (1986) and it requires estimating the following regressions.

$$Y = i_1 + cX + e_1 \tag{1}$$

$$Y = i_2 + c'X + bM + e_2 \tag{2}$$

$$M = i_3 + aX + e_3 \tag{3}$$

This type of mediation analysis is also called the product method as expressed below. Substituting Equation (3) into Equation (2) will yield Equation (2′).

$$Y = (i_3 + bi_2) + (a \times b + c')X + (be_2 + e_3) \tag{2'}$$

where $i_3 + bi_2 = i_1$, $a \times b + c' = c$ and $be_2 + e_3 = e_1$.

In this system, *c* or $a \times b + c'$ shows the total effect of *X* on *Y* and *c*′ shows direct effect of *X*. Therefore, *a*×*b* is the indirect effect of *X* on *Y* mediated by *M*. Baron and Kenny (1986) focused on non-zero coefficients instead of statistically significantly different from zero (Kline, 2015). Sobel (1982) defined an approximate test statistic that is distributed asymptotically normal is one of the approaches to test whether the product of *a* and *b* is statistically different from zero. Sobel's *z*-score is defined below with the variances of the sampling distributions of $\hat{a}$ and $\hat{b}$ OLS estimators in which the denominator shows the standard error of $\hat{a} \times \hat{b}$.

$$z_c = \frac{\hat{a} \times \hat{b}}{\sqrt{\hat{b}^2 \hat{\sigma}_a^2 + \hat{a}^2 \hat{\sigma}_b^2}}$$

There are possible three cases related to the measurement error: (i) error in *X* variable, (ii) error in the mediator (*M*) and (iii) errors in both *M* and *X*. We theoretically show that the naive OLS estimators (estimators with measurement error) under these three circumstances and our theoretical inferences are summarized as[4]:

- **Situation I: Error in *X*** $\left(\sigma_{u_1}^2 > 0 \,;\, \sigma_{u_2}^2 = 0\right)$

[4] Source: Authors' inferences.

$$c_{naive} = \lambda_1 . c_{true}$$

$$c'_{naive} = \frac{1}{1+\delta_1} c'_{true}$$

$$b_{naive} = b_{true} \cdot \frac{1}{1+\delta_1} + \frac{\delta_1}{1+\delta_1} \cdot \frac{\sigma_{YM}}{\sigma_M^2}$$

$$a_{naive} = \lambda_1 . a_{true}$$

- **Situation II: Error in *M*** $\left(\sigma_{u_1}^2 = 0 \; ; \; \sigma_{u_2}^2 > 0\right)$

$$c_{naive} = c_{true}$$

$$c'_{naive} = c'_{true} \cdot \frac{1}{1+\delta_2} + \frac{\delta_2}{1+\delta_2} \cdot \frac{\sigma_{YX}}{\sigma_X^2}$$

$$b_{naive} = b_{true} \cdot \frac{1}{1+\delta_2}$$

$$a_{naive} = a_{true}$$

- **Situation III: Errors in *X* and *M*** $\left(\sigma_{u_1}^2 > 0 \; ; \; \sigma_{u_2}^2 > 0\right)$

$$c_{naive} = \lambda_1 . c_{true}$$

$$c'_{naive} = c'_{true} \cdot \frac{1}{1+\delta_1+\delta_2+\left(1-\rho^2\right)\delta_1\delta_2} + \frac{\lambda_1 \delta_2 \left(1-\rho^2\right)}{1+\delta_2-\rho^2\left(\lambda_1+\delta_2\right)} \cdot \frac{\sigma_{YX}}{\sigma_X^2}$$

$$b_{naive} = b_{true} \cdot \frac{1}{1+\delta_1+\delta_2+\left(1-\rho^2\right)\delta_1\delta_2} + \frac{\lambda_2 \delta_1 \left(1-\rho^2\right)}{1+\delta_1-\rho^2\left(\lambda_2+\delta_1\right)} \cdot \frac{\sigma_{YX}}{\sigma_X^2}$$

$$a_{naive} = \lambda_1 . a_{true}$$

where ρ is the Pearson correlation coefficient between *X* and *M*. Besides, the definitions of the reliability ratios are given in Table 1.

Table 1. The definitions of the reliability ratios.

$\lambda_1 = \frac{\sigma_X^2}{\sigma_X^2 + \sigma_{u_1}^2}$	$\delta_1 = \frac{\sigma_{u_1}^2}{\left(1-\rho^2\right)\sigma_X^2}$
$\lambda_2 = \frac{\sigma_M^2}{\sigma_M^2 + \sigma_{u_2}^2}$	$\delta_2 = \frac{\sigma_{u_2}^2}{\left(1-\rho^2\right)\sigma_M^2}$

The interpretations of theoretical inferences for all three situations:

i. The coefficients c and a are affected only in Case I and III. Besides, these effects are correspond to the basic reliability ratio so-called attenuation factor, meaning that, the coefficients c and a are attenuated toward zero presence of the measurement error.

ii. The situation is different for the coefficients c' and b because their inferences are more complicated than the others. It is difficult to explore the effects of measurement error from Equation (2) which is the linear regression model with two explanatory variables and one of them is the mediator. Therefore, the inferences for this equation is related the correlation coefficient between *X* and *M*. This situation is required the new reliability ratios, δ_1 and δ_2.

iii. According to Table 1, these new reliability ratios cannot display any attenuation on the coefficients. These reliability ratios can influence both the decreasing and increasing direction of the true coefficients. In which direction the measurement error affects the parameter estimations depend on the variances of the explanatory variables, their measurement error variances and the correlation coefficient between these two variables.

3. The Simulation-Extrapolation (SIMEX) Method

Simulation-Extrapolation method is a parameter estimation technique based on simulation proposed by Cook and Stefanski (1994). With this method, the following linear regression model is estimated under the assumption of the explanatory variable *X* has a measurement error expressed by *U*.

$$Y = \alpha + \beta X + \varepsilon \qquad (4)$$

$$W = X + U \qquad (5)$$

As it is well-known, the OLS estimate of β is biased and inconsistent. The SIMEX method corrects the biasedness of the estimates. According to Carrol et al. (2006), the idea behind this method is to determine the effect of measurement error with the help of simulation and to eliminate this effect on the parameter estimates. The SIMEX method is the most effective and common method among the others since it does not require to know or to estimate the variance of measurement error. It should be noted that this method is defined under the classical measurement error definition that is given in Equation (5). There are studies applying this method different fields such as Kuchenoff et al. (2006).

The SIMEX method is applied in two steps named as simulation and extrapolation. The simulation step creates a new data set depending on the ordered values, $0 = \zeta_1 < \zeta_2 < \ldots < \zeta_M$, having increasing measurement error variance in addition to the observed data set. This variance is defined as

$$W_{b,i}(\zeta) = W_i + \zeta^{1/2} \sigma Z_{b,i} \qquad (6)$$

where $\left\{Z_{b,i}\right\}_{i=1}^{n}$ is a random variable generated form a normal distribution and *b* is the number of iteration ($b = 1, 2, \ldots, B$). First of all, each $\left\{Z_{b,i}\right\}_{i=1}^{n}$ random variable is used to estimate $\hat{\beta}_b(\zeta_m)$

coefficients via OLS estimator. Then, the scatter plot of each of ζ_m versus each of $\bar{\beta}(\zeta_m)$ is drawn by using the mean value obtained as

$$\bar{\beta}(\zeta_m) = \sum_b \hat{\beta}_b(\zeta_m) / B. \tag{7}$$

In the extrapolation step, the estimation is attained by the data with no measurement error. The scatter plot obtained in the first step is used to fit an approximate function such as rational extrapolant, quadratic extrapolant functions. In this recent paper, we used rational extrapolant function following the literature. The mean function given in Equation (7) is asymptotically shown as,

$$E\left(\hat{\beta}_m \mid \zeta\right) = f(\zeta) = \frac{\beta\sigma_x^2}{\sigma_x^2 + (1+\zeta)\sigma_u^2}, \ \zeta \geq 0 \tag{8}$$

where the extrapolation is applied to −1 point, $f(\zeta = -1)$, via the rational extrapolation function. The estimated coefficient is expressed by $\hat{\beta}_{SIMEX} = f(-1)$ and called as the SIMEX estimator (Cook and Stefanski, 1994; Carroll et al., 2006).

4. The Simulation Results

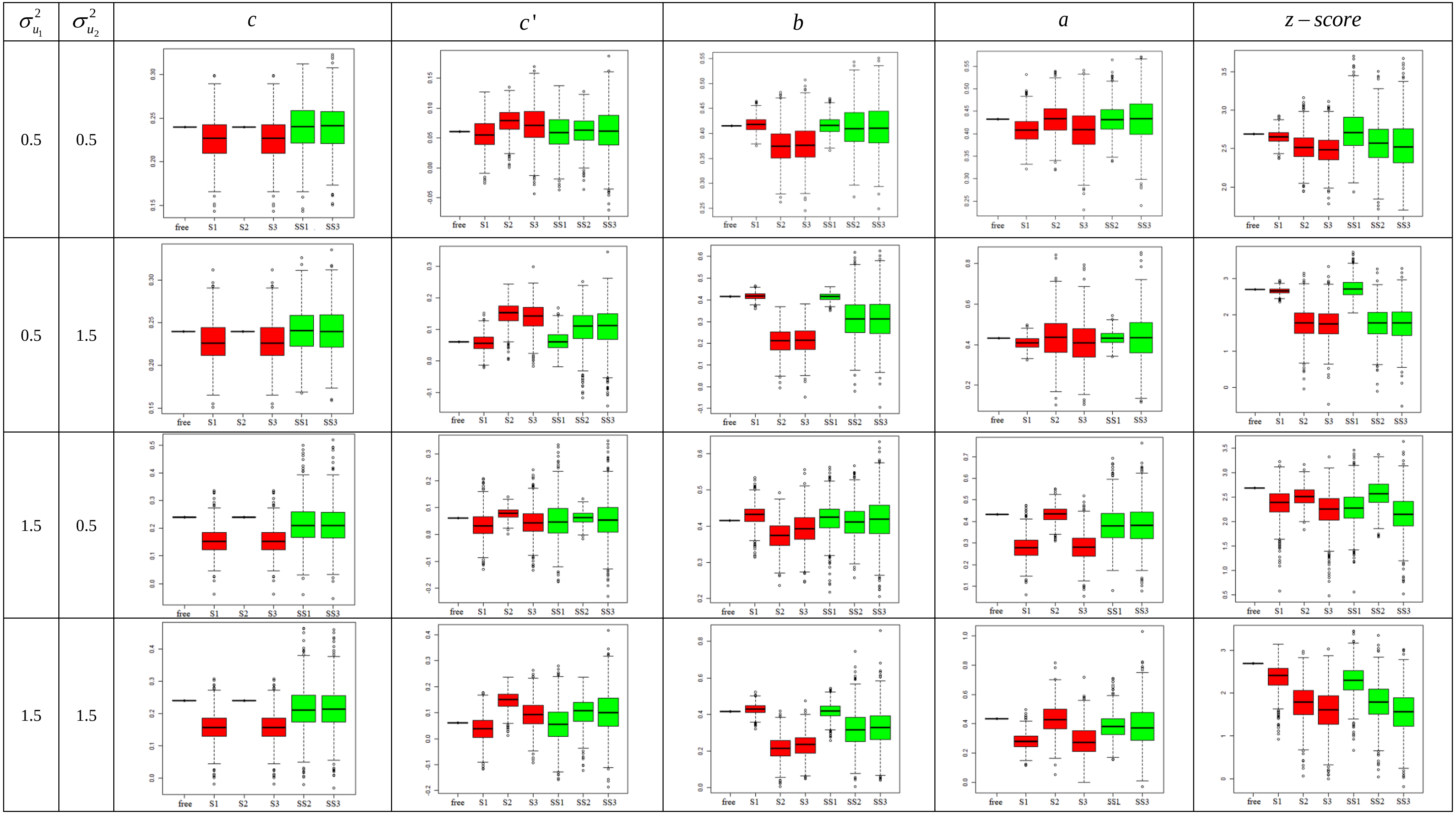
All the simulation analyses are run by R-Project programing language and used procedure is given below:

i. The variables *Y*, *X* and *M* are generated from a multiple multivariate normal distribution.
ii. Three different levels of linear relationship between the mediator and the variable *X* are considered since the mediation analysis requires relation between these two variables. The levels are named as small (30%), medium (50%) and large (70%).
iii. Sample sizes are chosen as 50, 100 and 500.
iv. Standard deviations of measurement errors for *X* and *M* are taken values in a sequence 0.5 to 5.5 in an increasing order by 0.5. In each loop 10000 repetitions are run.
v. The parameter estimations, *p*-values of estimates and Sobel's test statistics are calculated for both error-free and error-prone variables in three situations separately.

In this research, the results are given only under the conditions of n = 100 and medium-level relationship between *X* and *M* because of the similar results. The Figure 1 illustrates the estimates of the mediating system coefficients via OLS and SIMEX under the three situations. Table 2 and 3 demonstrate the naive and the SIMEX estimation results beyond the parameters of the mediating system and the Sobel's *z*-score under the different measurement error levels. According to these tables and figure, the most remarkable and important results are summarized as follows:

i. The estimate of the coefficient *c* is affected and attenuated towards zero from only measurement error in *X* not in *M*. Therefore, the estimate of *c* is appeared to be the same for Free and Case II; and Case I and Case III in all graphs. The only difference among these graphs is that the estimate of *c* decreases when measurement error in *X* increases.
ii. The results obtained for the coefficient c' under these three situations are more complicated as expected since c' belongs to Equation (2) including *M*. This coefficient is estimated downwards in Case I than the other cases. The measurement error in *M* is more dominated for c' and this

Figure 1. Box plots of the estimates of *c*, *c'*, *b*, *a* and Sobel's *z-score* under the different level of variances of measurement error and the SIMEX estimates (Red ones show error-prone estimates, green ones show SIMEX estimates).



coefficient is estimated upwards in Case II. The bias of the parameter estimation is similar to the Case II where the level of measurement error in *M* is higher than in *X* for Case III. The estimate of c' does not reach a balance even in the case of the same level of measurement error in both *X* and *M* (too high or too low). In summary, the total and direct effects that are c and c' respectively are both masked by measurement error even if variables have very small-level of measurement error.

iii. The Sobel’s test for mediation effect is required to estimate *a* and *b* coefficients as well as their standard errors. The coefficient *a* is slope parameter of the simple linear regression model (Eq. 1). Therefore, it should be mentioned that when *X* is mismeasured, its estimation (*a*) is attenuated towards zero like the estimate of *c*. In addition to the measurement error in *X*, the measurement error in *M* increases the biasedness. The raise in the measurement error in *M* could be suppressed the measurement error in *X*.

iv. The estimation of *b* appears to be one of the most affected one from the measurement error. Measurement error in mediator is attenuated the estimations towards zero remarkably because of *b* is the coefficient of the mediator. Against, the estimate of *b* under the measurement error in *X* approximates the error-free estimate of *b*. In summary, again, even if the variables have a very small variance of measurement error, error-free estimators are masked especially in case 3s.

v. Depend on the coefficient estimations, Sobel’s *z*-scores are attenuated towards zero either the measurement error in *X* or *M*. The error-prone Sobel’s *z*-scores cause the higher probability of the wrong decision for detecting the mediating effect.

vi. The SIMEX method produces the estimates close to true estimates of four parameters of the mediating system. Besides, for only the same and small variances of the measurement error, the SIMEX estimators can achieve the true *z*-score. Another result is that the SIMEX method generally underestimates the z-scores especially when the measurement error variances high.

vii. The SIMEX estimations for Sobel’s test can result in misleading inferences about the effect of the mediator. Increases in measurement error can increase the risk of making wrong decision in the applications.

Table 2. Naive and SIMEX estimations for mediating system under the different measurement error levels on S1, S2 and S3. (Error-free: c = 0.24; c' = 0.0605; b= 0.4151; a = 0.4325)

	$\sigma^2_{u_1}$	$\sigma^2_{u_2}$	c_{naive}	c_{SIMEX}	c'_{naive}	c'_{SIMEX}	b_{naive}	b_{SIMEX}	a_{naive}	a_{SIMEX}
S1	0.5	0.5	0.2264 (0.0006)	0.2401 (0.0007)	0.0558 (0.0007)	0.0597 (0.0009)	0.4178 (0.0002)	0.4156 (0.0003)	0.4085 (0.0009)	0.4329 (0.0011)
		1.5	0.2272 (0.0006)	0.2405 (0.0007)	0.0566 (0.0006)	0.0615 (0.0008)	0.4173 (0.0002)	0.4145 (0.0003)	0.4089 (0.0008)	0.4332 (0.0011)
		3.5	0.2248 (0.0006)	0.2382 (0.0007)	0.0541 (0.0007)	0.0583 (0.0009)	0.4188 (0.0002)	0.4164 (0.0003)	0.4076 (0.0008)	0.4316 (0.0010)
		5.5	0.2252 (0.0005)	0.2384 (0.0007)	0.0551 (0.0006)	0.0594 (0.0008)	0.4138 (0.0002)	0.4158 (0.0003)	0.4069 (0.0008)	0.4307 (0.0011)
	1.5	0.5	0.1553 (0.0023)	0.2131 (0.0047)	0.0351 (0.0025)	0.0509 (0.0053)	0.4299 (0.0009)	0.4208 (0.0019)	0.2801 (0.0031)	0.3841 (0.0070)
		1.5	0.1576 (0.0022)	0.2155 (0.0045)	0.0382 (0.0024)	0.0549 (0.0049)	0.4281 (0.0008)	0.4185 (0.0018)	0.2794 (0.0032)	0.3828 (0.0069)
		3.5	0.1556 (0.0024)	0.2135 (0.0051)	0.0349 (0.0008)	0.0499 (0.0054)	0.4299 (0.0009)	0.4212 (0.0019)	0.2813 (0.0032)	0.3856 (0.0070)
		5.5	0.1568 (0.0026)	0.2148 (0.0054)	0.0372 (0.0028)	0.0535 (0.0059)	0.4287 (0.0009)	0.4194 (0.0021)	0.2795 (0.0035)	0.3839 (0.0078)
	3.5	0.5	0.0579 (0.0019)	0.0975 (0.0056)	0.0108 (0.0017)	0.0186 (0.0051)	0.4440 (0.0007)	0.4394 (0.0021)	0.1063 (0.0026)	0.1798 (0.0080)
		1.5	0.0571 (0.0019)	0.0956 (0.0055)	0.0100 (0.0015)	0.0169 (0.0046)	0.4443 (0.0007)	0.4401 (0.0020)	0.1064 (0.0031)	0.1787 (0.0093)
		3.5	0.0583 (0.0018)	0.0975 (0.0053)	0.0106 (0.0015)	0.0177 (0.0047)	0.4441 (0.0006)	0.4402 (0.0019)	0.1078 (0.0026)	0.1805 (0.0079)
		5.5	0.0594 (0.0019)	0.0996 (0.0059)	0.0119 (0.0016)	0.0206 (0.0049)	0.4430 (0.0006)	0.4381 (0.0019)	0.1072 (0.0028)	0.1804 (0.0085)
	5.5	0.5	0.0341 (0.0013)	0.0594 (0.0042)	0.0075 (0.0010)	0.0129 (0.0033)	0.4456 (0.0005)	0.4423 (0.0016)	0.0600 (0.0019)	0.1048 (0.0063)
		1.5	0.0362 (0.0012)	0.0635 (0.0038)	0.0083 (0.0009)	0.0148 (0.0030)	0.4447 (0.0005)	0.4408 (0.0017)	0.0629 (0.0017)	0.1098 (0.0056)
		3.5	0.0341 (0.0013)	0.0594 (0.0041)	0.0072 (0.0010)	0.0126 (0.0034)	0.4457 (0.0006)	0.4425 (0.0018)	0.0605 (0.0018)	0.1059 (0.0058)
		5.5	0.0338 (0.0013)	0.0589 (0.0042)	0.0068 (0.0010)	0.0121 (0.0033)	0.4456 (0.0006)	0.4423 (0.0018)	0.0609 (0.0018)	0.1059 (0.0058)
S2	0.5	0.5	-	-	0.0778 (0.0004)	0.0617 (0.0006)	0.3754 (0.0013)	0.4132 (0.0018)	-	-
		1.5	-	-	0.1481 (0.0012)	0.1033 (0.0029)	0.2117 (0.0037)	0.3155 (0.0094)	-	-
		3.5	-	-	0.2114 (0.0008)	0.1908 (0.0025)	0.0654 (0.0021)	0.1132 (0.0067)	-	-
		5.5	-	-	0.2250 (0.0005)	0.2136 (0.0017)	0.0364 (0.0012)	0.0644 (0.0038)	-	-
	1.5	0.5	-	-	0.0780 (0.0005)	0.0619 (0.0006)	0.3746 (0.0015)	0.4122 (0.0021)	-	-
		1.5	-	-	0.1479 (0.0012)	0.1028 (0.0029)	0.2146 (0.0039)	0.3195 (0.0099)	-	-
		3.5	-	-	0.2108 (0.0008)	0.1894 (0.0024)	0.0653 (0.0022)	0.1134 (0.0068)	-	-
		5.5	-	-	0.2238 (0.0006)	0.2113 (0.0021)	0.0359 (0.0013)	0.0636 (0.0042)	-	-
	3.5	0.5	-	-	0.0773 (0.0005)	0.0609 (0.0006)	0.3774 (0.0015)	0.4148 (0.0019)	-	-
		1.5	-	-	0.1487 (0.0012)	0.1036 (0.0029)	0.2108 (0.0038)	0.3149 (0.0099)	-	-
		3.5	-	-	0.2115 (0.0008)	0.1902 (0.0024)	0.0661 (0.0022)	0.1146 (0.0069)	-	-

		5.5	-	-	0.2249 (0.0005)	0.2132 (0.0018)	0.0358 (0.0013)	0.0636 (0.0043)	-	-
	5.5	0.5	-	-	0.0764 (0.0005)	0.0599 (0.0006)	0.3765 (0.0015)	0.4142 (0.0021)	-	-
		1.5	-	-	0.1476 (0.0014)	0.1021 (0.0034)	0.2138 (0.0039)	0.3195 (0.0101)	-	-
		3.5	-	-	0.2123 (0.0008)	0.1919 (0.0024)	0.0647 (0.0021)	0.1126 (0.0066)	-	-
		5.5	-	-	0.2245 (0.0005)	0.2123 (0.0017)	0.0350 (0.0014)	0.0623 (0.0046)	-	-
S3	0.5	0.5	0.2264 (0.0006)	0.2400 (0.0007)	0.0722 (0.0010)	0.0623 (0.0014)	0.3784 (0.0015)	0.4127 (0.0021)	0.4082 (0.0021)	0.4325 (0.0025)
		1.5	0.2272 (0.0006)	0.2405 (0.0007)	0.1394 (0.0018)	0.1061 (0.0038)	0.2146 (0.0038)	0.3160 (0.0097)	0.4107 (0.0115)	0.4348 (0.0131)
		3.5	0.2248 (0.0005)	0.2382 (0.0006)	0.1973 (0.0014)	0.1897 (0.0032)	0.0671 (0.0022)	0.1147 (0.0067)	0.4046 (0.0587)	0.4286 (0.0664)
		5.5	0.2252 (0.0005)	0.2386 (0.0006)	0.2107 (0.0010)	0.2118 (0.0022)	0.0373 (0.0011)	0.0653 (0.0037)	0.3950 (0.1210)	0.4203 (0.1369)
	1.5	0.5	0.1553 (0.0022)	0.2129 (0.0048)	0.0453 (0.0027)	0.0552 (0.0058)	0.3919 (0.0021)	0.4172 (0.0037)	0.2812 (0.0042)	0.3853 (0.0088)
		1.5	0.1576 (0.0021)	0.2158 (0.0044)	0.0935 (0.0029)	0.1032 (0.0070)	0.2324 (0.0039)	0.3298 (0.0103)	0.2776 (0.0105)	0.3800 (0.0210)
		3.5	0.1556 (0.0024)	0.2137 (0.0050)	0.1342 (0.0029)	0.1722 (0.0071)	0.0739 (0.0023)	0.1234 (0.0073)	0.2939 (0.0445)	0.4034 (0.0857)
		5.5	0.1568 (0.0026)	0.2146 (0.0053)	0.1449 (0.0029)	0.1915 (0.0067)	0.0405 (0.0013)	0.0693 (0.0043)	0.2913 (0.0847)	0.3999 (0.1665)
	3.5	0.5	0.0578 (0.0019)	0.0974 (0.0057)	0.0141 (0.0017)	0.0214 (0.0053)	0.4111 (0.0016)	0.4391 (0.0034)	0.1069 (0.0030)	0.1805 (0.0091)
		1.5	0.0577 (0.0019)	0.0961 (0.0056)	0.0312 (0.0018)	0.0440 (0.0056)	0.2492 (0.0034)	0.3538 (0.0095)	0.1064 (0.0058)	0.1781 (0.0168)
		3.5	0.0583 (0.0018)	0.0978 (0.0053)	0.0492 (0.0019)	0.0778 (0.0063)	0.0842 (0.0023)	0.1425 (0.0074)	0.1073 (0.0191)	0.1813 (0.0559)
		5.5	0.0594 (0.0019)	0.1001 (0.0058)	0.0551 (0.0021)	0.0908 (0.0065)	0.0462 (0.0014)	0.0804 (0.0046)	0.1041 (0.0388)	0.1742 (0.1135)
	5.5	0.5	0.0340 (0.0013)	0.0596 (0.0041)	0.0093 (0.0011)	0.0148 (0.0035)	0.4119 (0.0015)	0.4407 (0.0032)	0.0606 (0.0022)	0.1059 (0.0069)
		1.5	0.0363 (0.0012)	0.0634 (0.0039)	0.0201 (0.0011)	0.0301 (0.0037)	0.2545 (0.0034)	0.3628 (0.0091)	0.0638 (0.0034)	0.1109 (0.0107)
		3.5	0.0341 (0.0013)	0.0598 (0.0041)	0.0293 (0.0014)	0.0489 (0.0045)	0.0850 (0.0021)	0.1443 (0.0069)	0.0589 (0.0108)	0.1030 (0.0336)
		5.5	0.0337 (0.0013)	0.0593 (0.0041)	0.0311 (0.0013)	0.0532 (0.0043)	0.0470 (0.0014)	0.0817 (0.0046)	0.0627 (0.0199)	0.1105 (0.0631)

The values in the parenthesis are shown the standard errors of the parameter estimations.

Table 3. Naive and SIMEX estimations for Sobel *z*-score under the different measurement error levels on S1, S2 and S3. (Error-free: *z*-score = 2.688)

$\sigma^2_{u_1}$	$\sigma^2_{u_2}$	S1 z_{naive}	S1 z_{SIMEX}	S2 z_{naive}	S2 z_{SIMEX}	S3 z_{naive}	S3 z_{SIMEX}
0.5	0.5	2.652 (0.0072)	2.735 (0.0749)	2.520 (0.0332)	2.571 (0.0714)	2.487 (0.0389)	2.546 (0.1055)
	1.5	2.652 (0.0066)	2.736 (0.0697)	1.7588 (0.1789)	1.7664 (0.1908)	1.7423 (0.1772)	1.760 (0.2152)
	3.5	2.655 (0.0075)	2.728 (0.0719)	0.7712 (0.3737)	0.7697 (0.3831)	0.7643 (0.3752)	0.7715 (0.4029)
	5.5	2.651 (0.0073)	2.724 (0.0737)	0.4582 (0.3534)	0.4619 (0.3674)	0.4589 (0.3466)	0.4636 (0.3732)
1.5	0.5	2.364 (0.0985)	2.2806 (0.1246)	2.519 (0.0399)	2.576 (0.0746)	2.234 (0.1329)	2.158 (0.1611)
	1.5	2.3598 (0.1015)	2.2878 (0.1241)	1.7664 (0.1816)	1.7799 (0.2012)	1.5905 (0.0541)	1.5513 (0.2482)
	3.5	2.3681 (0.1047)	2.2885 (0.1291)	0.7797 (0.3555)	0.7951 (0.3761)	0.7135 (0.3784)	0.6977 (0.4011)
	5.5	2.3546 (0.1113)	2.2780 (0.1342)	0.4760 (0.3937)	0.4823 (0.4078)	0.4301 (0.3698)	0.4103 (0.3875)
3.5	0.5	1.5620 (0.4673)	1.5380 (0.4463)	2.5270 (0.0372)	2.5790 (0.0739)	1.4950 (0.4751)	1.4670 (0.4530)
	1.5	1.5640 (0.5579)	1.5430 (0.5358)	1.7499 (0.1806)	1.7616 (0.2001)	1.0780 (0.4993)	1.0607 (0.5073)
	3.5	1.5950 (0.4645)	1.5640 (0.4375)	0.7701 (0.3645)	0.7774 (0.3728)	0.4899 (0.4409)	0.4813 (0.4568)
	5.5	1.5770 (0.4954)	1.5500 (0.4712)	0.4418 (0.3752)	0.4453 (0.3799)	0.2562 (0.4281)	0.2369 (0.4441)
5.5	0.5	1.1974 (0.6948)	1.1844 (0.6947)	2.5300 (0.0383)	2.5810 (0.0718)	1.1494 (0.6863)	1.1389 (0.6829)
	1.5	1.2602 (0.6005)	1.2442 (0.5845)	1.7663 (0.1932)	1.7796 (0.2064)	0.8926 (0.5745)	0.8817 (0.5814)
	3.5	1.2108 (0.6509)	1.2022 (0.6389)	0.7498 (0.3806)	0.7586 (0.3944)	0.3465 (0.4824)	0.3311 (0.4918)
	5.5	1.2186 (0.6436)	1.1974 (0.6154)	0.4636 (0.3658)	0.4681 (0.3802)	0.2180 (0.4025)	0.2099 (0.4175)

The values in the parenthesis are shown the standard errors of the parameter estimations.

5. Conclusion

This paper investigates (1) the behaviors of the OLS estimators and the test statistics in the standard mediation analysis of Baron and Kenny (1986) when the mediator and/or the explanatory variable are measured with error, (2) whether the SIMEX method corrects the measurement error from the both mediating system parameters and the Sobel's test statistics when the true *M* and/or *X* are unobservable.

The Monte Carlo simulation results show that the measurement error in the mediator influences the mediation analysis more than the measurement error in regressor. Increasing the sample size helps to obtain correct results under measurement error in *M*, but it is not guaranteed to obtain the reliable statistical inferences. It should be also highlighted that increasing variance of the measurement error in *M* rises the effects of measurement error in the mediating analysis. The higher correlations between *X* and *M* alleviate the effects of measurement error in the mediation analysis while not for the weak correlation. In this instance, the SIMEX methodology can assist to produce the almost unbiased estimators for the parameters of the mediating system while not displaying same accuracy for Sobel's *z*-statistics. For why, the SIMEX methodology is based on the Jacknife method that causes the higher standard errors of the estimators.

According to the results of this paper, it is suggested that the researchers using mediation analysis should pay attention to the measurement error in regressors of the analysis. Besides, the SIMEX methodology is successful for the parameter estimations of the mediating system but SIMEX could improve for the calculation of the Sobel's *z*-scores presence of the measurement error.

References

Barron, R. M., & Kenny, D. A. (1986). The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations. Journal of Personality and Social Psychology. 51(6): 1173-1182.

Carroll, R. J., Ruppert, D., Stefanski, L. A., Crainiceanu, C. M. (2006). Measurement Error in Nonlinear Models. USA:Chapman&Hall/CRC.

Cook, J. R., and Stefanski, L. A. (1994). Simulation-Extrapolation Estimation in Parametric Measurement Error Models. Journal of the American Statistical Association, 89(428), 1314-1328.

Fritz, M. S., Kenny, D. A., MacKinnon, D. P. (2016). The Combined Effects of Measurement Error and Omitting Confounders in the Single-Mediator Model. Multivariate Behavior Research. 51(5): 681-697.

Judd, C. M., & Kenny, D. A. (1981). Process Analysis: Estimating Mediation in Evaluation Research. Evaluation Research, 5: 602-619.

Kline, R. B. (2015). The Mediation Myth. Basic and Applied Social Psychology. 37(4): 202-213, DOI: 10.1080/01973533.2015.1049349

Kuchenoff, H., Mwalili, S.M. and Lesaffre, E. (2006). A General Method for Dealing with Misclassification in Regression: The Misclassification SIMEX. Biometrics, 62, 85-96.

le Cessie, S., Debeij, J., Rosendaal, F. R., Cannegieter, S. C., Vandenbrouckea, J. P. (2012) Quantification of Bias in Direct Effects Estimates Due to Different Types of Measurement Error in the Mediator. Epidemiology. 23(4): 551-560.

Sobel, M. E. (982). Asymptotic Confidence Intervals for Indirect Effects in Structural Equation Models. Sociological Methodology. 113: 290-312.

VanderWeele, T. J., Valeri, L., Ogburn, E. L. (2012). The role of measurement error and misclassification in mediation analysis. Epidemiology. 23(4): 561-564.

Vargas, Y. G. W. (2015). Causal Inference Methods for Measurement Error and Mediation. Unpublished Ph. D. Thesis, The Johns Hopkins University, Baltimore-Maryland.