

Appendix C (Pornprasertmanit & Little, in press)

Supplementary Simulation Study Using the Target Paper's Guideline

In this simulation, X is an explanatory variable, Y is a response variable, and Z is a covariate. X and Z were independent ($\rho_{XZ} = 0$). In this simulation, we only studied the directional dependency determined by skewness under four simulation conditions:

1. Skewness of explanatory variable: Strongly Negative ($\gamma_1 \approx -2$), Negative ($\gamma_1 \approx -1$), Normal ($\gamma_1 = 0$), Positive ($\gamma_1 \approx 1$), and Strongly Positive ($\gamma_1 \approx 2$).
2. Skewness of covariate: Strongly Negative, Negative, Normal, Positive, and Strongly Positive.
3. Correlation (standardized regression coefficient) between the response variable and the explanatory variable: .2 and .5.
4. Correlation (standardized regression coefficient) between the response variable and the covariate: 0 (baseline model), .2, .5, and .8.

To create different levels of skewness, we used a chi-squared distribution. Chi-squared distribution has a higher level of skewness if its degree of freedom is low. In the strongly negative and the strongly positive skewness condition we used the chi-square distribution with 2 degrees of freedom. In the negative and the positive skewness condition we used the chi-square distribution with 10 degrees of freedom. In the normal condition we used a standard normal distribution. We examined only the sample size of 1,000. Ten thousand replications were run per condition.

We used the target paper's guideline for determining directional dependency. Recall that if the skewness values of both X and Y is not significant, the directional dependency is undetermined. If the skewness of X or Y is significant, then directional dependency can be

determined. If the skewness of X is greater than Y , X would be suggested as an explanatory variable. If the skewness of Y is greater than X , Y would be suggested as an explanatory variable. We used X as the explanatory variable in this simulation study, so our results should suggest X as the explanatory variable. We used g_1 as the skewness index. The Wald statistic is used to examine the significance of an individual skewness with standard error (SE_{g_1} ; Joanes & Gill, 1998) of

$$SE_{g_1} = \sqrt{\frac{6n(n-1)}{(n-2)(n+1)(n+3)}} \quad (C1)$$

Where n is sample size.

The results of the simulation study are shown in Tables C1 and C2, which show the results when the correlations between X and Y are .2 and .5, respectively. The rows of both tables are divided into four panels that represent the correlation between Y and Z (0, .2, .5, and .8, respectively). There are three possible results from the directional dependency test: X is the explanatory variable, Y is the explanatory variable, or the directional dependency is undetermined. Each cell represents the proportion that the directional dependency showed each possible result across 10,000 replications in a condition.

The first panel of rows of both tables is the baseline model without the covariate effect. When the explanatory variable is skewed, the directional dependency can be determined 100% correctly. On the other hand, if the explanatory variable is normal, 90% of replications were undetermined, 5% of replications suggested X as an explanatory variable, and 5% of replications suggested Y as an explanatory variable. The similar pattern of result is generalized to the second panel of rows of both tables when the effect of the covariate is small (i.e., the correlation between covariate and response variable is .2).

The third and fourth panels of rows of both tables represent the conditions where the effect of the covariate is medium ($\rho = .5$) and large ($\rho = .8$), respectively. In the normal explanatory variable and skewed covariate, more than 50% of replications indicated that Y was the explanatory variable, which is wrong. If the effect of the covariate is large (i.e., the fourth layer), a mildly skewed explanatory variable in one direction (e.g., negatively skewed) is not robust enough to an extremely skewed covariate in another direction (e.g., positively skewed). As a result, the skewness of the covariate makes the response variable more skewed than the explanatory variable, which leads to an error in determining the direction of the effect.

From this simple simulation, we can conclude that the test of directional dependency is robust to the small effects of unobserved explanatory variables. When the covariate effect is medium or large, the result cannot be trusted because the skewness of a variable can result from either itself or the covariate. Thus, directional dependency can only be trusted if the existence of a nontrivial, unobserved explanatory variable can be ruled out. In other words, the correlations of the covariate with both variables used in directional dependency should not exceed .2. Although we do not investigate the effect of the correlated covariate, as shown in Equations A23 and A31, a small correlation between the covariate and the explanatory variable should not make a strong difference because Equation A31 is just an addition and a subtraction to the similar terms from Equation A23.

From this simulation, Type I error can be examined when X is normal and the correlation between X and Z is 0 (the third row of the first panel). In this situation, the decision should be that the directional dependency is undetermined (see the top panel of Table 1). The proportion that the test suggested “undetermined” is .90. That is, the overall Type I error is .10, which is systematically greater than the nominal level of .05. Our proposed method is better in controlling

overall Type I error.

Supplementary Materials for

Pornprasertmanit, S., & Little, T. D. (in press). Determining directional dependency in causal associations. *International Journal of Behavioral Development*.

Table C1

The Simulation Results of Directional Dependency if a Covariate Exists and the Correlation

Between Explanatory and Response Variables Are .2 ($\rho_{XY} = .2$)

Explanatory Variable Skewness	Covariate Skewness														
	SNeg			Neg			Norm			Pos			SPos		
	X	Y	U	X	Y	U	X	Y	U	X	Y	U	X	Y	U
Covariate Correlation = 0 (No Covariate Effect)															
SNeg	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00
Neg	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00
Norm	.05	.05	.90	.05	.05	.90	.05	.05	.90	.05	.05	.90	.05	.05	.90
Pos	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00
SPos	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00
Covariate Correlation = .2															
SNeg	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00
Neg	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00
Norm	.05	.06	.90	.05	.05	.90	.05	.05	.90	.05	.05	.90	.05	.06	.89
Pos	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00
SPos	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00
Covariate Correlation = .5															
SNeg	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00
Neg	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00
Norm	.02	.84	.15	.03	.51	.46	.05	.05	.90	.03	.51	.46	.02	.84	.14
Pos	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00
SPos	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00
Covariate Correlation = .8															
SNeg	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00
Neg	.87	.13	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	.89	.11	.00
Norm	.00	1.00	.00	.00	1.00	.00	.05	.05	.90	.00	1.00	.00	.00	1.00	.00
Pos	.88	.12	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	.87	.13	.00
SPos	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00

Note. X = Proportion of replications suggesting X as the explanatory variable, which is true. Y =

Proportion suggesting Y as the explanatory variable, which is false. U = Proportion of

undetermined directional dependency replications. SNeg = Strongly negative skewed. Neg =

Negative skewed. Norm = Normal distribution. Pos = Positive skewed. SPos = Strongly positive

skewed. The bold numbers represent the conditions that are not 100% accurate (i.e., when the

proportion of replications suggesting X as explanatory variable is less than 1.0).

Table C2

The Simulation Results of Directional Dependency if a Covariate Exists and the Correlation

Between Explanatory and Response Variables Are .5 ($\rho_{XY} = .5$)

Explanatory Variable Skewness	Covariate Skewness														
	SNeg			Neg			Norm			Pos			SPos		
	X	Y	U	X	Y	U	X	Y	U	X	Y	U	X	Y	U
Covariate Correlation = 0 (No Covariate Effect)															
SNeg	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00
Neg	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00
Norm	.05	.05	.91	.05	.05	.90	.05	.05	.90	.05	.05	.90	.05	.05	.91
Pos	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00
SPos	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00
Covariate Correlation = .2															
SNeg	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00
Neg	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00
Norm	.05	.06	.89	.04	.05	.91	.05	.05	.91	.05	.05	.90	.05	.06	.90
Pos	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00
SPos	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00
Covariate Correlation = .5															
SNeg	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00
Neg	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00
Norm	.01	.84	.15	.03	.51	.46	.05	.05	.90	.03	.51	.47	.01	.85	.14
Pos	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00
SPos	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00
Covariate Correlation = .8															
SNeg	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00
Neg	.67	.33	.00	1.00	.01	.00	1.00	.00	.00	1.00	.00	.00	.96	.04	.00
Norm	.00	1.00	.00	.00	1.00	.00	.05	.05	.90	.00	1.00	.00	.00	1.00	.00
Pos	.96	.04	.00	1.00	.00	.00	1.00	.00	.00	1.00	.01	.00	.67	.33	.00
SPos	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.01	.00

Note. X = Proportion of simulations suggesting X as the explanatory variable, which is true. Y =

Proportion suggesting Y as the explanatory variable, which is false. U = Proportion of

undetermined directional dependency replications. SNeg = Strongly negative skewed. Neg =

Negative skewed. Norm = Normal distribution. Pos = Positive skewed. SPos = Strongly positive

skewed. The bold numbers represent the conditions that are not 100% accurate(i.e., when the

proportion of replications suggesting X as explanatory variable is less than 1.0).