simulation summary

Joakim Wallmark

11/2/2019

Simulation study

Goal

Parametric estimators for natural direct- and indirect effects are implemented in the sensmediation package. The goal is to test the methods implemented in this package. We will investigate and compare bias from model misspecification (interaction term missing in outcome model) with bias from mediator-outcome confounding through simulations.

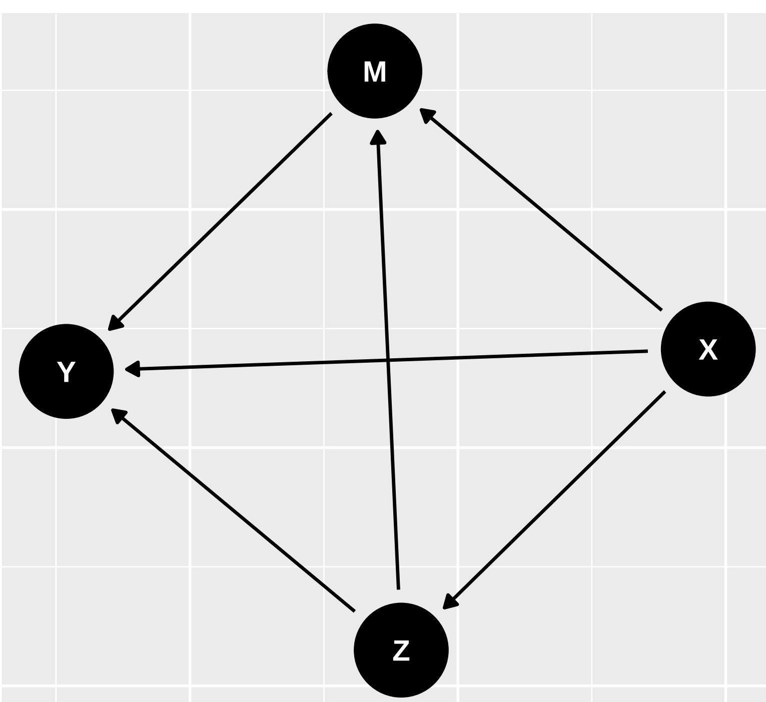
Simulation scenarios

Simulations of bias were done with model misspecification and confounding in separate comparable scenarios. 100000 iterations with a sample size of 1000 was used for each scenario. Mediator and outcome were made continous and similar simulations using binary mediator/outcome will be done later.

Table 1: Variables used for simulations

variable	type	true model
X(additional covariate)	continous	$X \sim gamma(8, 4.5)$
Z(exposure)	binary	$Z = I(Z* > 0)$ where $Z* \sim U_0 + U_1X + N(0, 1)$
M(mediator)	continous	$M \sim B_0 + B_1Z + B_2X + N(0, 1)$
Y(outcome)	continous	$Y \sim \theta_0 + \theta_1Z + \theta_2M + \theta_3ZM + \theta_4X + N(0, 1)$

DAG showing data generating process



Simulation scenarios (model misspecification)

Parameter values used for simulations:

- $U_0 = -0.4$
- $U_1 = 0.01$
- $B_0 = 3$
- $B_1 = 2$
- $B_2 = 0.05$
- $\theta_0 = 5$
- $\theta_1 = 1$
- $\theta_2 = 0.5$
- $\theta_3 =$ varied[-0.5, 0.5]
- $\theta_4 = 0.05$

Estimated mediator model was set to the correct one. Estimated outcome model was linear but misspecified without ZM interaction: $Y \sim Z + M + X$.

Simulation scenarios (mediator-outcome confounding)

Parameter values used for simulations:

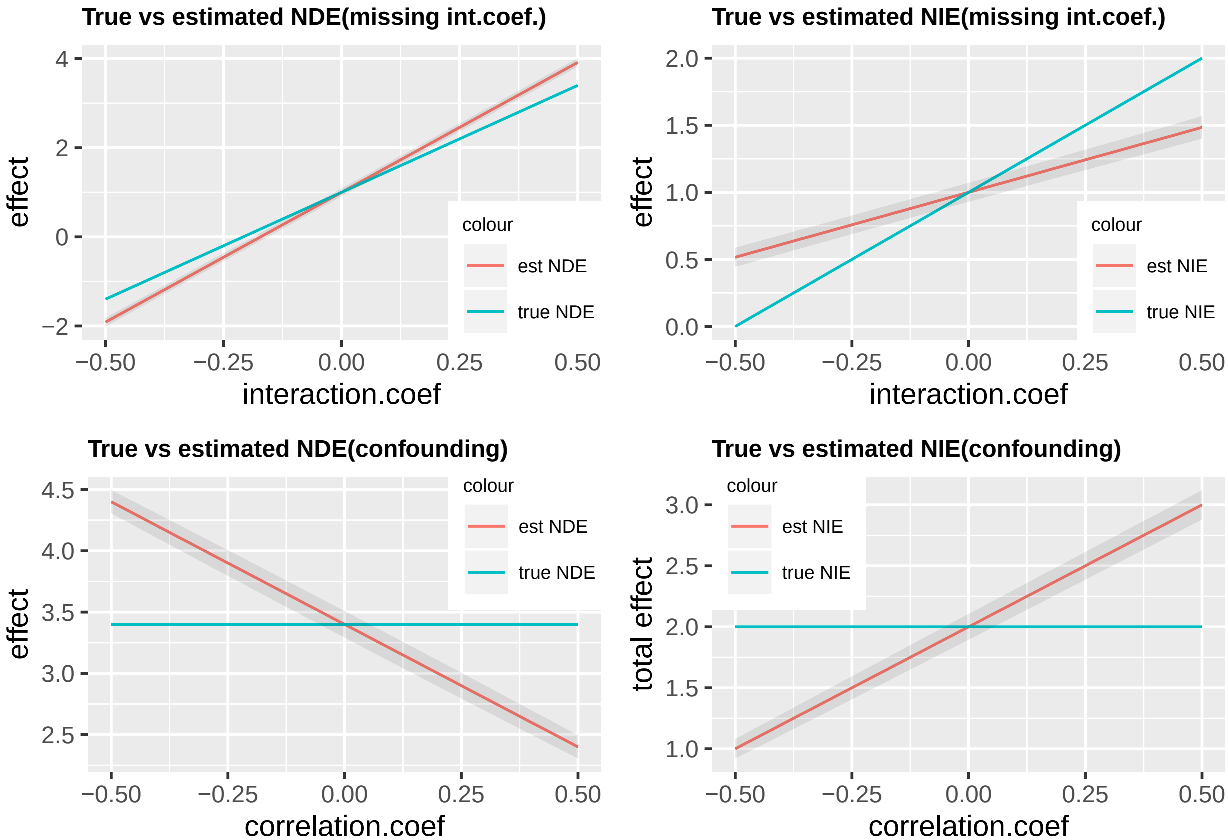
$corr(\omega, \epsilon)$ =varied (-0.5,0.5)

- $U_0 = -0.4$
- $U_1 = 0.01$
- $B_0 = 3$
- $B_1 = 2$
- $B_2 = 0.05$
- $\theta_0 = 5$
- $\theta_1 = 1$
- $\theta_2 = 0.5$
- $\theta_3 = 0.5$
- $\theta_4 = 0.05$

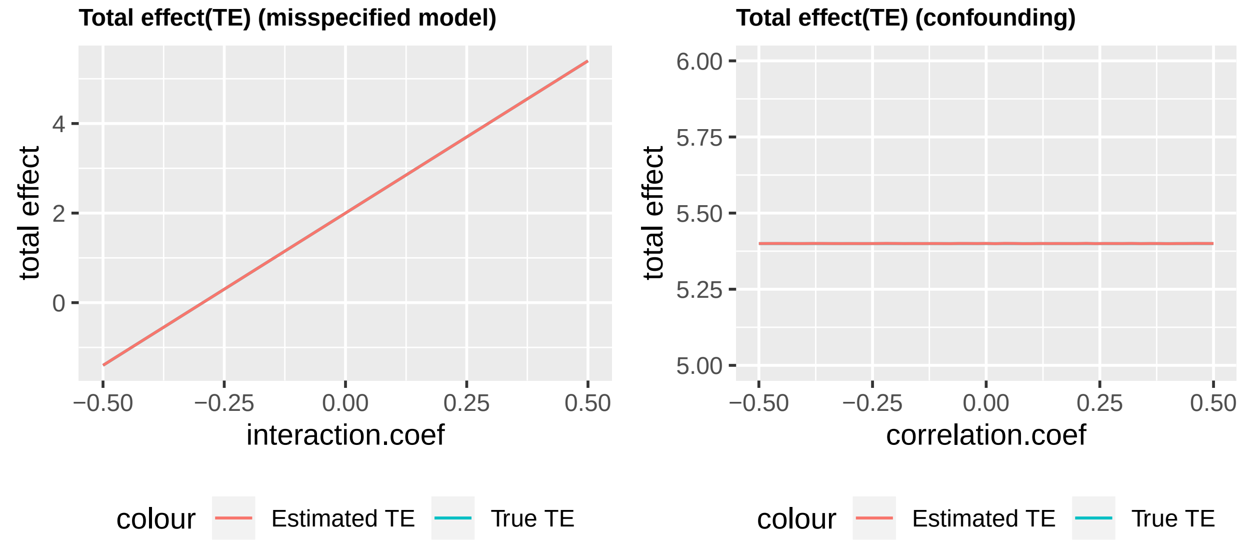
Estimated mediator model was set to the correct one. Bias was induced by changing error term correlation (between ω and ϵ).

Results

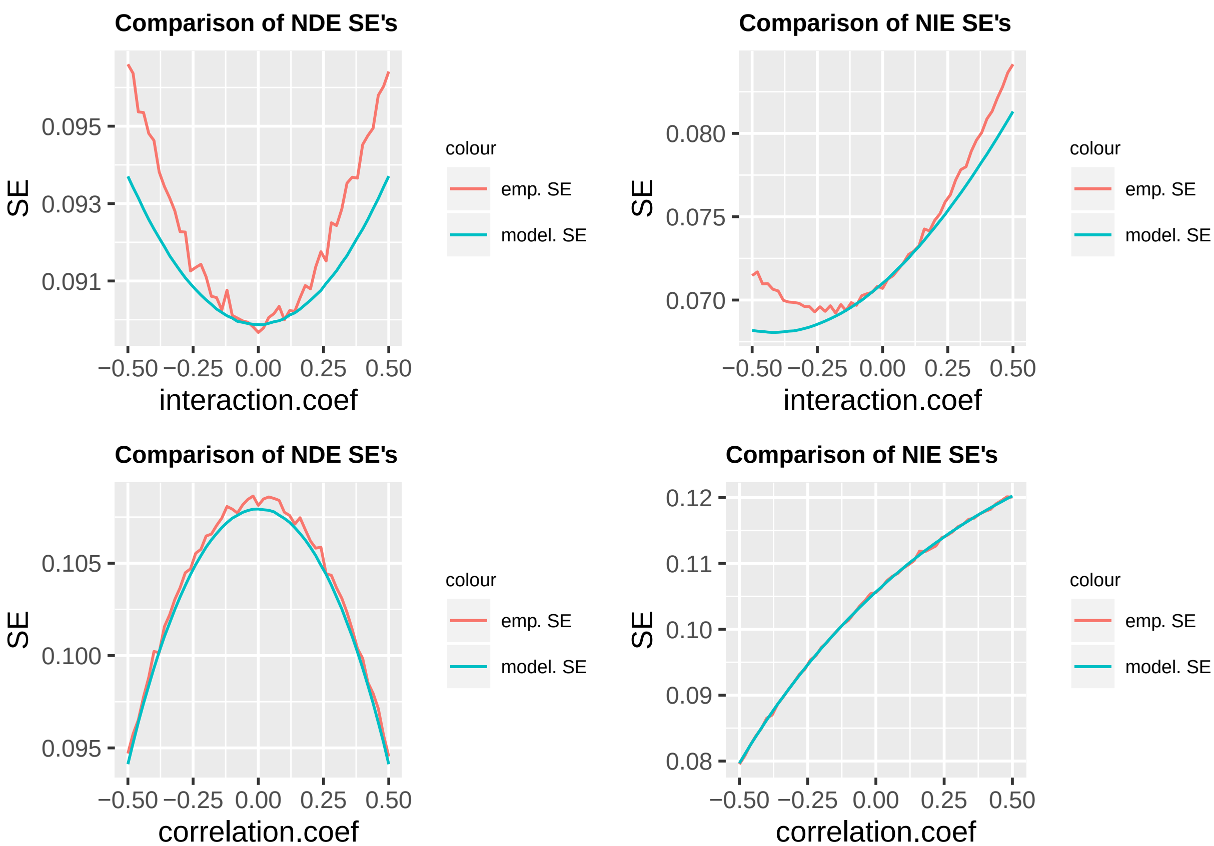
Bias of NDE and NIE changes linearly as we change confounding or interaction strength in the respective scenarios:



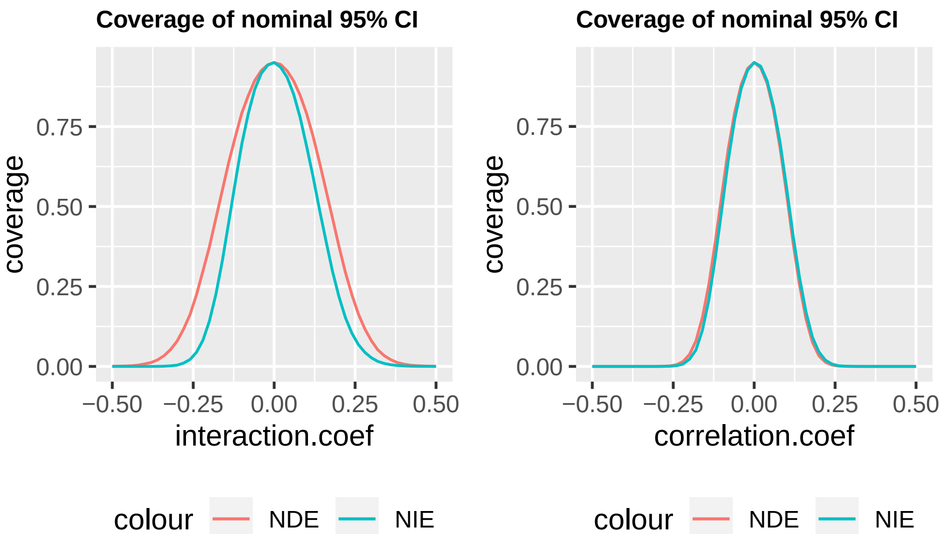
Total effect is unbiased because the NDE and NIE biases will cancel eachother out:



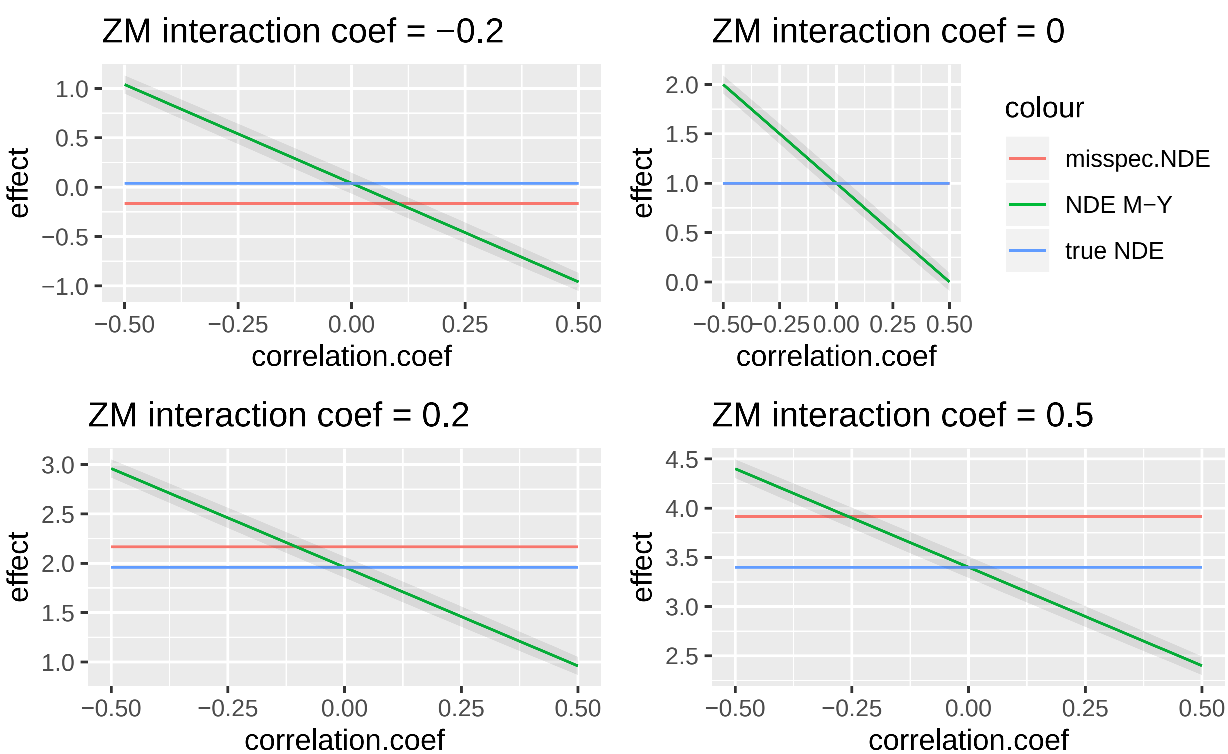
When missing the interaction term in the outcome model, the empirical data shows that we tend to underestimate the SE's of our estimates as the interaction strength increases. In the case of unobserved confounding, the SE's are still pretty accurate:



As expected coverage decreases as the interaction get stronger when we misspecify our outcome model. The same happens for larger confounding. Since a big n(1000) was used for simulations we get bad coverage even when the bias is low:



Comparison between confounding and model misspecification bias for different interaction strengths between Z and M. Example: In this scenario, when the true interaction coefficient is 0.5, we need a correlation between Z and M of about -0.25 to achieve the same amount of bias from confounding as we would get from missing the interaction term in our outcome model:



Because of linearity and same total effect, the correlation needed to match the bias of an outcome model missing an interaction term with a coefficient size of 0.5 is -0.25 for for the NIE too: