Simulation details

This document serves as a guide for replicating the simulation exercise in the main paper. All relevant code can be found at https://github.com/keithbarnatchez/ME-Causal-Review.

Data generation

The simulated data is comprised of

- an outcome Y
- a covariate X with error-prone measurements W
- an error-free covariate Z
- a binary treatment T

The data generating process is comprised of 4 models: 1) an outcome model, 2) a treatment model, 3) a covariate model, and 4) a measurement model.

Outcome model

$$Y = \beta_0 + \beta_X X + \beta_Z Z + \beta_T T + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2_\varepsilon)$$

Treatment model

$$T|X, Z \sim \text{Bernoulli}(p), \quad \text{logit}(p) = \alpha_0 + \alpha_X X + \alpha_Z Z$$

Covariate model

$$(X, Z) \sim \text{MVN}\left(\mathbf{0}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right)$$

We also consider the scenario where there is an instrument V for X, in which case

$$(X, Z, V) \sim \text{MVN}\left(\mathbf{0}, \begin{bmatrix} 1 & \rho & \psi \\ \rho & 1 & 0 \\ \psi & 0 & 1 \end{bmatrix}\right)$$

By construction, note that $\varepsilon \perp V$.

Measurement model

We assume a classical, non-differential measurement error process:

$$W = X + U, \quad U \sim N(0, \sigma^2_U)$$