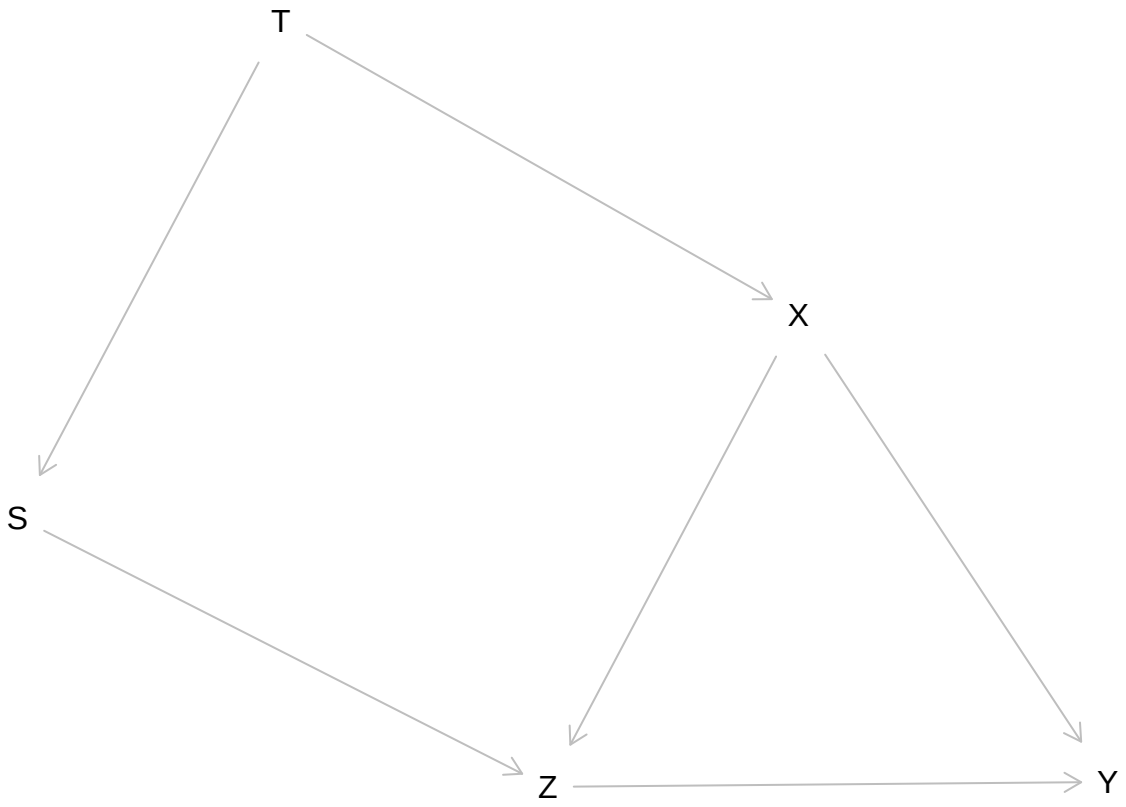
Simulation Study to estimate NDE and NIE

2025-12-01

Causal dag fig 2(a)



Data generation from the causal dag

```
true_nde <- 2.0
true_nie <- 1.5
generate_data <- function(n) {
  T <- rnorm(n)
  X <- rbinom(n, 1, plogis(0.5 * T))
  S <- 0.5 * T + rnorm(n)
  Z <- 1 + 1 * X + 0.5 * S + rnorm(n)
  Y <- 2 + true_nde * X + true_nie * Z + rnorm(n)

  data.frame(Y = Y, X = X, Z = Z, S = S, T = T)
}
```

Estimator NDE

```
estimate_NDE <- function(data, x = 1, x_star = 0) {
  # Fit models
  model_Y <- lm(Y ~ X + Z, data = data)
  model_Z <- lm(Z ~ X + S, data = data)

  n <- nrow(data)
  nde_sum <- 0

  for (i in 1:n) {
    s_i <- data$S[i]

    # Predict Z given X=x_star, S=s_i
    z_pred <- predict(model_Z, newdata = data.frame(X = x_star, S = s_i))

    # E[Y|X=x, Z=z_pred] - E[Y|X=x_star, Z=z_pred]
    y_x <- predict(model_Y, newdata = data.frame(X = x, Z = z_pred))
    y_xstar <- predict(model_Y, newdata = data.frame(X = x_star, Z = z_pred))

    nde_sum <- nde_sum + (y_x - y_xstar)
  }

  return(nde_sum / n)
}
```

Estimator NIE

```
estimate_NIE <- function(data, x = 1, x_star = 0) {
  # Fit models
  model_Y <- lm(Y ~ X + Z, data = data)
  model_Z <- lm(Z ~ X + S, data = data)

  n <- nrow(data)
  nie_sum <- 0

  for (i in 1:n) {
    s_i <- data$S[i]

    # Predict Z given X=x, S=s_i
    z_x <- predict(model_Z, newdata = data.frame(X = x, S = s_i))

    # Predict Z given X=x_star, S=s_i
    z_xstar <- predict(model_Z, newdata = data.frame(X = x_star, S = s_i))

    # E[Y|X=x_star, Z=z_x] - E[Y|X=x_star, Z=z_xstar]
    y_zx <- predict(model_Y, newdata = data.frame(X = x_star, Z = z_x))
    y_zxstar <- predict(model_Y, newdata = data.frame(X = x_star, Z = z_xstar))

    nie_sum <- nie_sum + (y_zx - y_zxstar)
  }

  return(nie_sum / n)
```

```
}
```

Run the simulations

```
sample_sizes <- c(100, 500, 1000, 2000, 5000, 10000)
n_sims <- 50

results <- data.frame()

for (n in sample_sizes) {
  cat("Generating data for n =", n, "...\n")

  nde_estimates <- replicate(n_sims, {
    data <- generate_data(n)
    estimate_NDE(data)
  })

  nie_estimates <- replicate(n_sims, {
    data <- generate_data(n)
    estimate_NIE(data)
  })

  results <- rbind(results, data.frame(
    n = n,
    NDE_mean = mean(nde_estimates),
    NDE_bias = mean(nde_estimates) - true_nde,
    NIE_mean = mean(nie_estimates),
    NIE_bias = mean(nie_estimates) - true_nie
  ))
}
```

```
## Generating data for n = 100 ...
## Generating data for n = 500 ...
## Generating data for n = 1000 ...
## Generating data for n = 2000 ...
## Generating data for n = 5000 ...
## Generating data for n = 10000 ...
```

```
print(results)
```

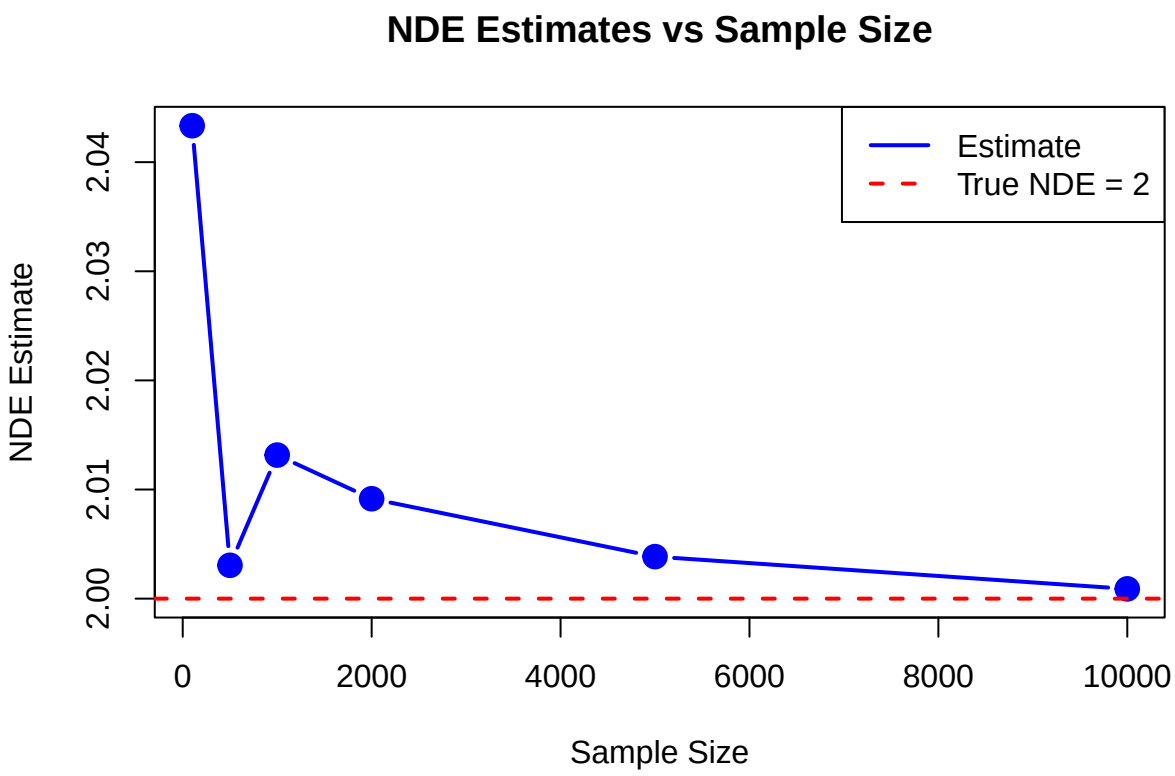
```
##       n NDE_mean    NDE_bias NIE_mean      NIE_bias
## 1   100 2.043338 0.043338339 1.486562 -0.0134376552
## 2   500 2.003054 0.003053908 1.532511  0.0325106601
## 3  1000 2.013155 0.013155482 1.490295 -0.0097046933
## 4  2000 2.009153 0.009153471 1.503851  0.0038508528
## 5  5000 2.003849 0.003848887 1.492249 -0.0077511241
## 6 10000 2.000894 0.000894082 1.500653  0.0006527569
```

Check if the estimators are asymptotically unbiased

NDE Estimates

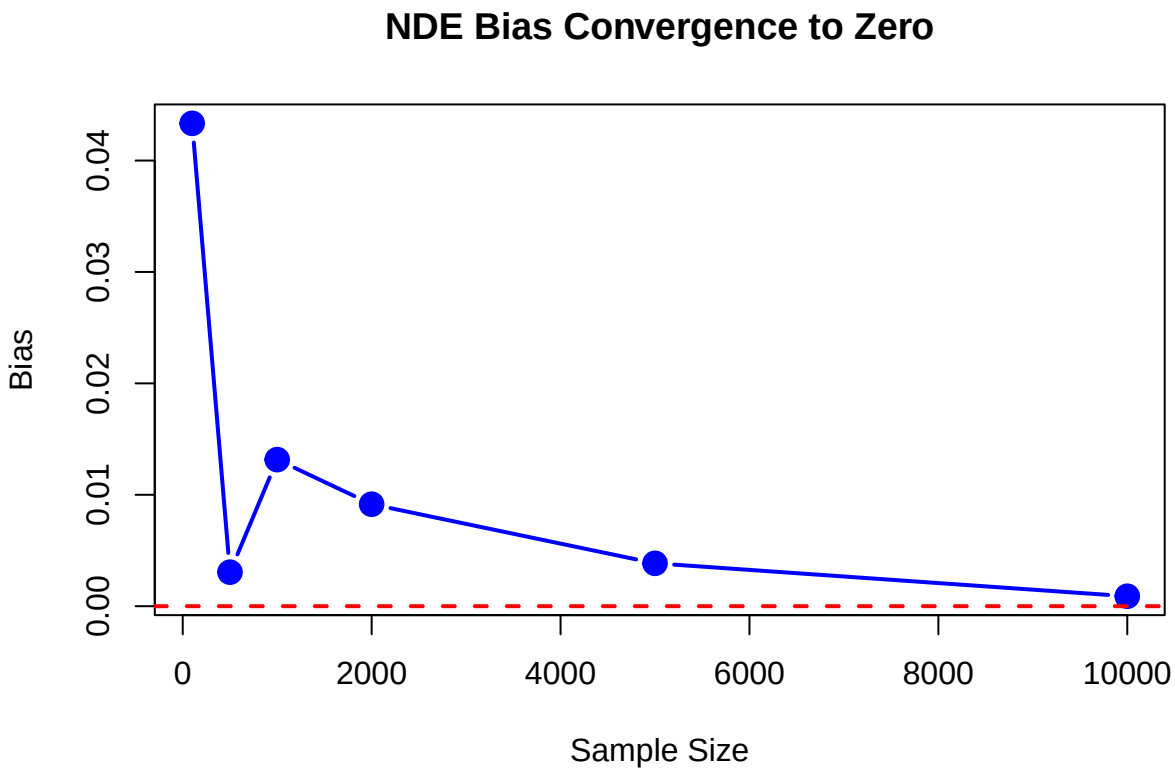
```
plot(results$n, results$NDE_mean,
     type = "b", pch = 19, col = "blue", lwd = 2, cex = 1.5,
```

```
     xlab = "Sample Size", ylab = "NDE Estimate",
     main = "NDE Estimates vs Sample Size",
     ylim = range(c(results$NDE_mean, 2)))
abline(h = 2, col = "red", lty = 2, lwd = 2)
legend("topright", legend = c("Estimate", "True NDE = 2"),
       col = c("blue", "red"), lty = c(1, 2), lwd = 2)
```



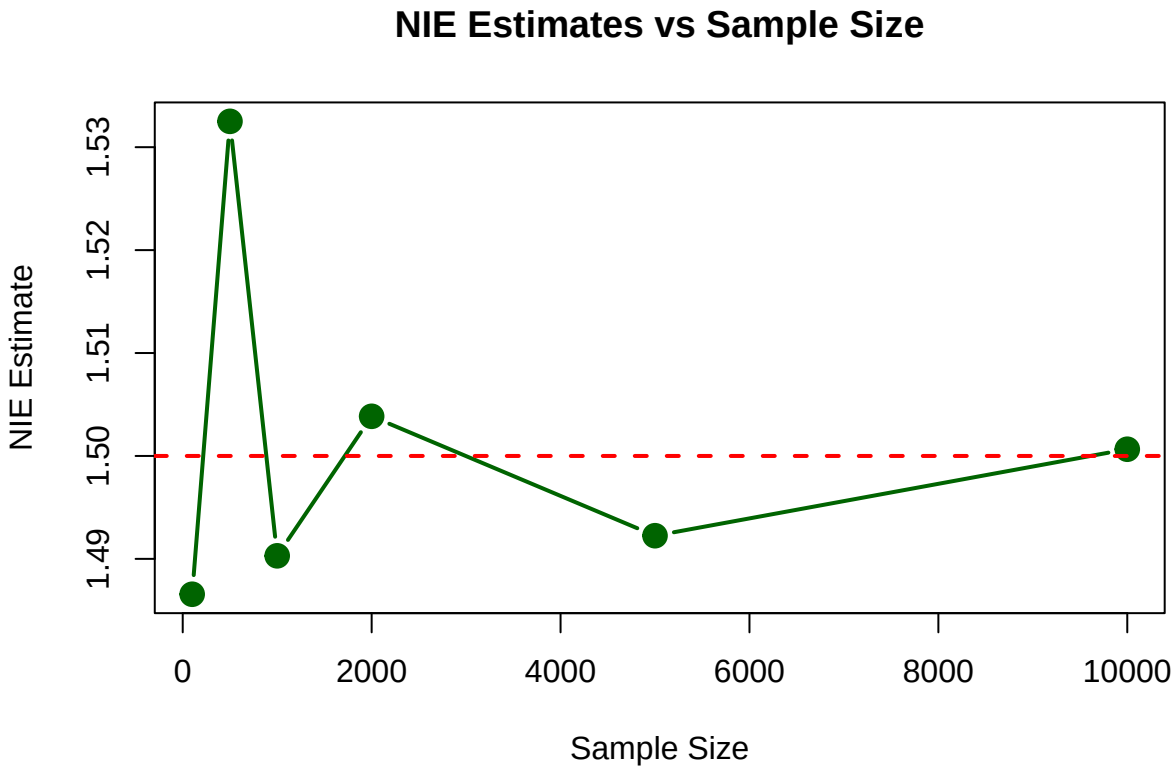
NDE Bias

```
plot(results$n, results$NDE_bias,
     type = "b", pch = 19, col = "blue", lwd = 2, cex = 1.5,
     xlab = "Sample Size", ylab = "Bias",
     main = "NDE Bias Convergence to Zero")
abline(h = 0, col = "red", lty = 2, lwd = 2)
```



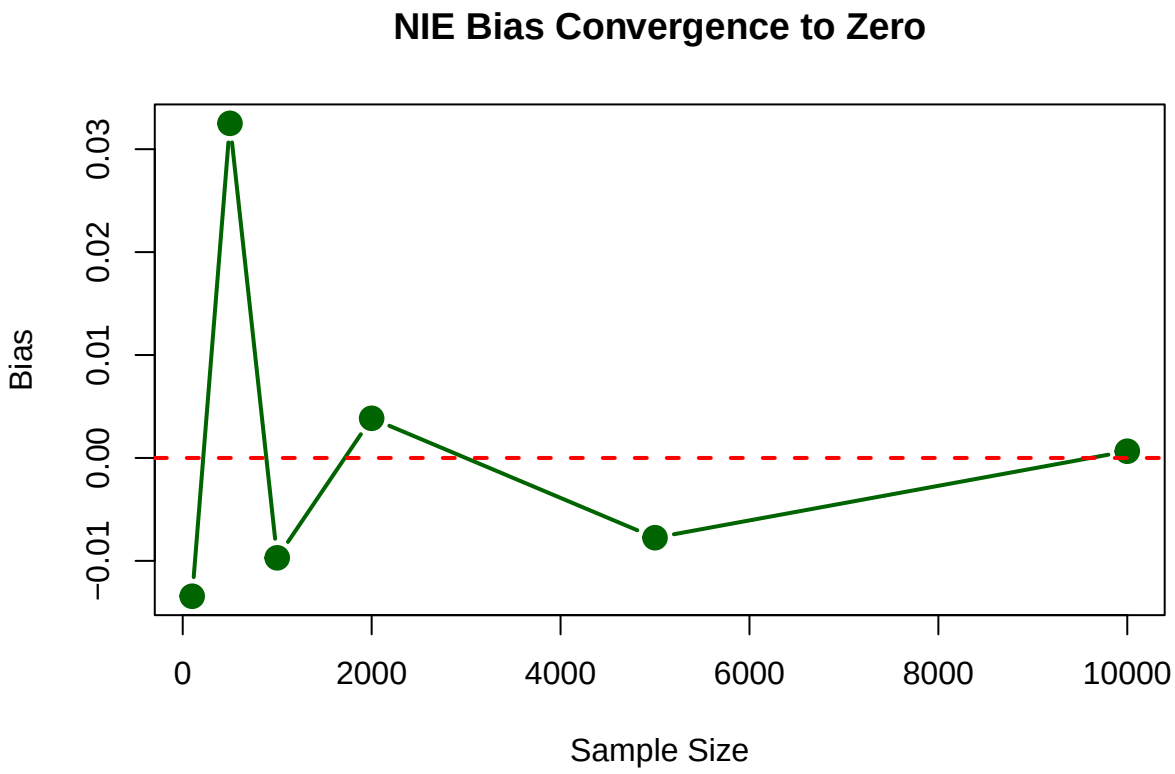
NIE Estimates

```
plot(results$n, results$NIE_mean,
     type = "b", pch = 19, col = "darkgreen", lwd = 2, cex = 1.5,
     xlab = "Sample Size", ylab = "NIE Estimate",
     main = "NIE Estimates vs Sample Size",
     ylim = range(c(results$NIE_mean, 1.5)))
abline(h = 1.5, col = "red", lty = 2, lwd = 2)
```



NIE Bias

```
plot(results$n, results$NIE_bias,
     type = "b", pch = 19, col = "darkgreen", lwd = 2, cex = 1.5,
     xlab = "Sample Size", ylab = "Bias",
     main = "NIE Bias Convergence to Zero")
abline(h = 0, col = "red", lty = 2, lwd = 2)
```


NIE Bias Convergence to Zero
Bias
0.03
0.02
0.01
0.00
−0.01
0
2000
4000
6000
8000
10000
Sample Size