Causal Inference Using Graphs: Problem Set 1 Solution

1

a. $X \rightarrow Z_1 \rightarrow Z_3 \rightarrow Z_5 \rightarrow Y$
$X \rightarrow Z_2 \rightarrow Z_3 \rightarrow Z_5 \rightarrow Y$
$X \rightarrow Z_1 \rightarrow Z_3 \leftarrow Z_4 \rightarrow Y$
$X \rightarrow Z_2 \rightarrow Z_3 \leftarrow Z_4 \rightarrow Y$

b. $X \rightarrow Z_1 \rightarrow Z_3 \rightarrow Z_5 \rightarrow Y$
$X \rightarrow Z_2 \rightarrow Z_3 \rightarrow Z_5 \rightarrow Y$

c. Z_5

2

a.
```
# Set seed:
set.seed(123)

# Given quantities:
n <- 1000
beta1 <- 1
delta <- .8
rho <- .8

# Number of datasets:
S <- 10000

# Vectors for each estimate:
est1 <- rep(0,S)
est2 <- rep(0,S)

# Simulation:
for(s in 1:S){
  # Error terms nu & gamma:
  nu1 <- rnorm(n)
  nu2 <- rnorm(n)
  gamma1 <- rnorm(n)
  gamma2 <- rnorm(n)
  # x's:
  x1 <-  gamma1
```

```
  x2 <- delta * x1 + gamma2
  # Error terms epsilon:
  eps1 <- nu1
  eps2 <- rho * eps1 + nu2
  # y's:
  y1 <- beta1*x1 + eps1
  y2 <- beta1*x2 + eps2
  # Estimates of beta_1:
  est1[s] <- lm(y2 ~ x2)$coef[2]
  est2[s] <- lm(y2 ~ x2 + y1)$coef[2]
}

# Average of the 10,000 estimates:
mean(est1)
mean(est2)
# The simple regression y_2 ~ x_2 seems to provide an unbiased estimate

# Explore graphically:
par(mfrow=c(2,1))
hist(est1)
hist(est2)
# Confirms the intuition from the means above
```

b. Conditioning on the collider, y_{t-1}, opens the backdoor path between x_t and y_t, which was previously blocked by the collider itself. Hence, regressing y_t on x_t and y_{t-1} induces bias. If y_{t-1} does not have a causal effect on the outcome variable (irrelevant variable in the regression), the simulation results invalidate the claim made in most econometrics textbooks.