Problem Set 2 Solution: Causal Graphs

Problem 1 (2 points)

Consider the following two period TSCS model where we assume constant effects:

$$
\begin{aligned}
y_{i,2} &= \beta_0 + \beta_1 x_{i,2} + \beta_2 x_{i,1} + \beta_3 y_{i,1} + \epsilon_{i,2} \\
\epsilon_{i,2} &= \rho \epsilon_{i,1} + \nu_{i,2} \\
x_{i,2} &= \delta_1 x_{i,1} + \delta_2 y_{i,1} + \gamma_{i,2}
\end{aligned}
$$

You have the option of estimating the effect of $x_{i,2}$ on $y_{i,2}$ using the coefficient on $x_{i,2}$ from the following regressions:

1. $y_2 \sim x_2$
2. $y_2 \sim x_2 + y_1$
3. $y_2 \sim x_2 + x_1$
4. $y_2 \sim x_2 + x_1 + y_1$

Consider the situation where some of following may equal zero: β_2, β_3, ρ, δ_1, and δ_2.

a) Suppose that $\beta_2 = \beta_3 = \delta_2 = 0$. Which regression or regressions will produce a consistent estimator for β_1?

   1, 3, and 4 will be consistent.

b) Suppose that $\beta_2 = 0$ which regression or regressions will produce a consistent estimator for β_1?

   With $\beta_3 \neq 0$ and $\delta_2 \neq 0$, the lagged DV is a common cause of both $x_{i,2}$ and $y_{i,2}$, hence it must be included and only 4 will remain consistent.

Problem 2 (5 points)

Consider the following two period TSCS model where we assume constant effects:

$$y_{i,2} = \beta_0 + \beta_1 x_{i,2} + \beta_2 x_{i,1} + \beta_3 y_{i,1} + \epsilon_{i,2}$$
$$\epsilon_{i,2} = \rho_\epsilon \epsilon_{i,1} + \nu_{i,2}$$
$$x_{i,2} = \delta_1 x_{i,1} + \delta_2 y_{i,1} + \gamma_{i,2}$$
$$\gamma_{i,2} = \rho_\gamma \gamma_{i,1} + \omega_{i,2}$$

You have the option of estimating the effect of $x_{i,2}$ on $y_{i,2}$ using the coefficient on $x_{i,2}$ from the following regressions:

1. $y_2 \sim x_2$
2. $y_2 \sim x_2 + y_1$
3. $y_2 \sim x_2 + x_1$
4. $y_2 \sim x_2 + x_1 + y_1$

Consider the situation where some of following may equal zero: β_2, β_3, ρ_ϵ, δ_1, δ_2, and ρ_γ.

a) Suppose that $\beta_2 = \beta_3 = \rho_\epsilon = \delta_1 = \delta_2 = 0$ but that $\rho_\gamma \neq 0$. Which regression or regressions will produce a consistent estimator for β_1?

All four will be consistent.

b) Suppose that $\beta_2 = \beta_3 = \delta_1 = \delta_2 = 0$ but that $\rho_\gamma \neq 0$ and $\rho_\epsilon \neq 0$. Which regression or regressions will produce a consistent estimator for β_1?

Regressions 1, 3, and 4 will produce consistent estimates

c) Suppose that $\beta_2 = \delta_1 = \delta_2 = 0$ but that $\beta_3 \neq 0$, $\rho_\gamma \neq 0$ and $\rho_\epsilon \neq 0$. Which regression or regressions will produce a consistent estimator for β_1?

Regressions 3 and 4 will produce consistent estimates

d) Suppose that $\beta_1 = \beta_2 = \beta_3 = \delta_1 = \rho_\gamma = 0$ but that $\delta_2 \neq 0$ and $\rho_\epsilon \neq 0$. Will regression 1 or 3 produce more bias for an estimate of β_1?

Regression 3 will produce more bias due to bias amplification.

e) Suppose that $\beta_1 = \beta_2 = \beta_3 = \delta_1 = \rho_\gamma = 0$ but that $\delta_2 \neq 0$ and $\rho_\epsilon \neq 0$ and suppose you do not observe y_1 (i.e., regressions 2 and 4 are no longer options). Propose a consistent estimator for β_1.

Regress $y_2 \sim x_1$, regress $x_2 \sim x_1$, take the ratio of the slope coefficients (i.e., use the Wald IV estimator).

Problem 3 (3 points, plus bonus)

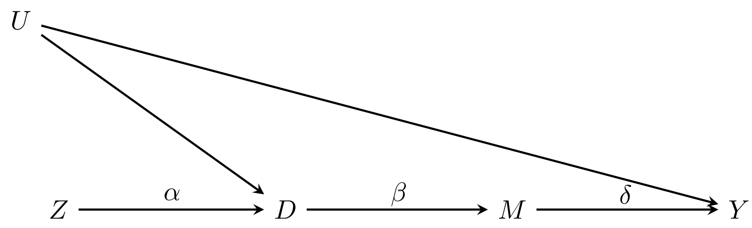

Figure 1: Linear Constant-Effects SEM. U is unobserved.

a) What is the effect of D on Y?

$\beta \cdot \delta$

b) Describe how you could estimate the effect of D on Y using the Wald estimator.

$lm(y \sim z)$ provides $\widehat{\alpha \cdot \beta \cdot \delta}$. $lm(d \sim z)$ provides $\widehat{\alpha}$. The ratio provides $\widehat{\beta \cdot \delta}$

c) Which regression would produce more bias for the effect of D on Y: $lm(y \sim d)$ or $lm(y \sim d + z)$? Why?

$lm(y \sim d + z)$ produces more bias because of amplification.

d) (bonus) Describe some additional ways you could estimate the effect of D on Y.

1. $lm(y \sim m + d)$ provides $\widehat{\delta}$. $lm(m \sim d)$ provides $\widehat{\beta}$. Take product.
2. $lm(m \sim z)$ provides $\widehat{\alpha \cdot \beta}$. $lm(d \sim z)$ provides $\widehat{\alpha}$. The ratio provides $\widehat{\beta}$. $lm(y \sim m + d)$ provides $\widehat{\delta}$. Take product.
3. $lm(y \sim z)$ provides $\widehat{\alpha \cdot \beta \cdot \delta}$. $lm(m \sim z)$ provides $\widehat{\alpha \cdot \beta}$. The ratio provides $\widehat{\delta}$. $lm(m \sim d)$ provides $\widehat{\beta}$. Take product.

Problem 4 (4 points)

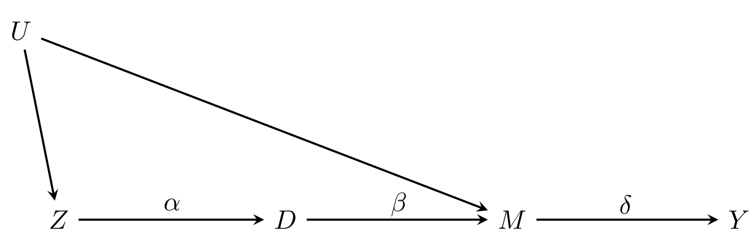

Figure 2: Linear Constant-Effects SEM. U is unobserved.

a) What is the effect of Z on Y?

$\alpha \cdot \beta \cdot \delta$

b) Describe one way you could consistently estimate the effect of Z on Y.

$lm(y \sim m)$ provides $\widehat{\delta}$. $lm(m \sim d + z)$ provides $\widehat{\beta}$. $lm(d \sim z)$ provides $\widehat{\alpha}$. The product provides $\widehat{\alpha \cdot \beta \cdot \delta}$

c) What is the effect of D on Y?

$\beta \cdot \delta$

d) Describe two ways you could consistently estimate the effect of D on Y.

1. $lm(y \sim d + z)$ provides $\widehat{\beta \cdot \delta}$
2. $lm(y \sim m)$ provides $\widehat{\delta}$. $lm(m \sim d + z)$ provides $\widehat{\beta}$. Take product.