Omitted Variable Bias Calculation

Full Model: TS ~ B0 + B1(CS) + B2(SES) + e1

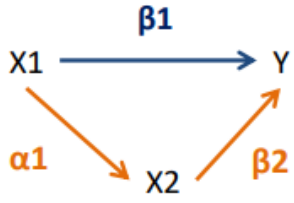
Naïve Model: TS ~ b0 + b1(CS) + e2

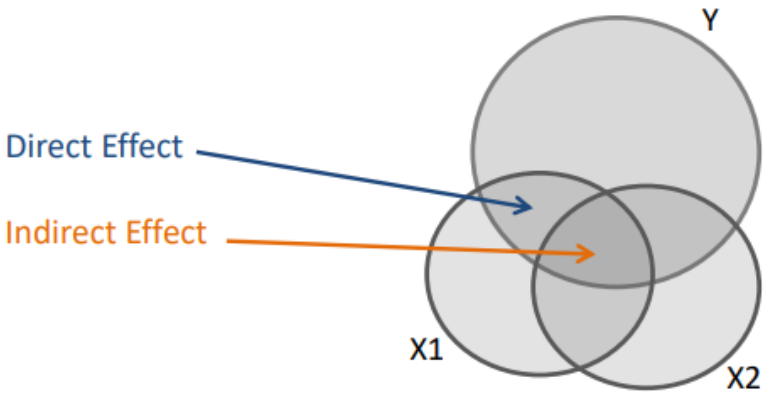
Policy variable: Classroom Size

Omitted variable: Socio-Economic Status

```
URL <- "https://raw.githubusercontent.com/DS4PS/cpp-523-fall-
2019/master/labs/class-size-seed-1234.csv"
dat <- read.csv( URL )

m.full <- lm( test ~ csize + ses, data=dat  )
summary( m.full )  # Model 4

m.naive <- lm( test ~ csize, data=dat  )
summary( m.naive)  # Model 1

m.auxiliary <- lm( ses ~ csize, data=dat )
summary( m.auxiliary )

# lm(formula = ses ~ csize, data = dat)
# Coefficients:
# ---------------------------------
#              Estimate Std. Error t value Pr(>|t|)
# (Intercept)  4.469458   0.009033   494.8   <2e-16 ***
# csize       -0.094876   0.000326  -291.0   <2e-16 ***
# ---------------------------------


# b1 = B1 + bias
# b1 - B1 = bias
b1 <- -4.22
B1 <- -2.67
b1 - B1

# bias = a1*B2
a1 <- -0.0949
B2 <- 16.34
a1*B2
```

	Dependent Variable: Test Scores				
	Model 1	Model 2	Model 3	Model 4	Model 5
	(1)	(2)	(3)	(4)	(5)
Classroom Size	-4.22***	-3.91***		-2.67	-2.22***
	(0.18)	(0.03)		(1.63)	(0.23)
Teacher Quality		55.01***	55.03***		55.01***
		(0.25)	(0.26)		(0.25)
Socio-Economic Status			40.94***	16.34	17.77***
			(0.27)	(17.10)	(2.40)
Intercept	738.34***	456.70***	272.91***	665.29***	377.26***
	(4.88)	(1.48)	(1.39)	(76.57)	(10.82)





$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon_1$$ (full regression)

$$X_2 = \alpha_0 + \alpha_1 X_1 + \varepsilon_2$$ (auxiliary regression)

$$bias = \beta_2 \alpha_1$$ (path diagram for X1→X2→Y)