Exercise 7

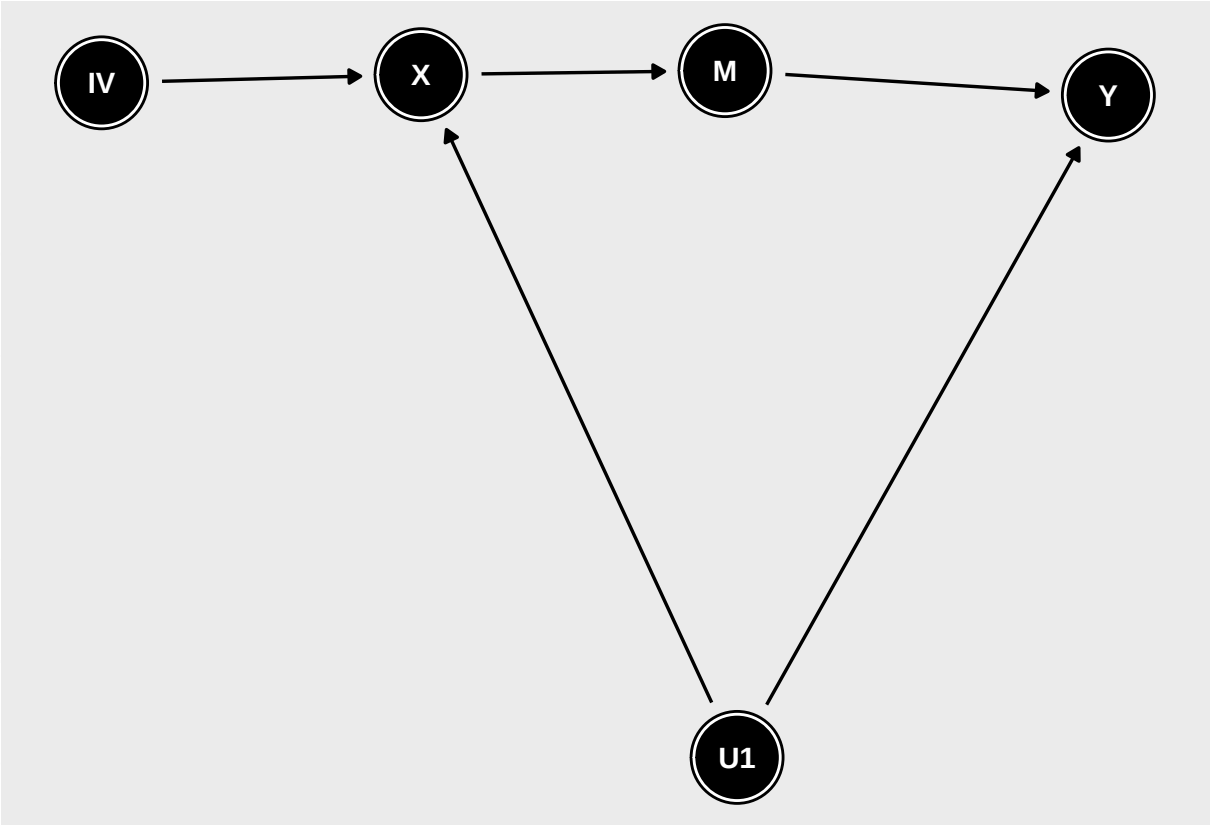
This seventh exercise is an add-on that quickly explores instrumental variables, and mechanisms. No special file needs to be downloaded. We are just simulating some data on the fly, and you can reproduce the analyses if you wish.

I am first simulating data from the following simple model, that contains both an IV, and a mechanism. Both of them can be used to recover the true effect, in the presence of unobserved confounding.

```
library(ggdag)
library(dagitty)
library(emmeans)
library(lavaan)
library(AER)

dag7 <- dagitty('dag {
bb="0,0,1,1"
IV [pos="0.116,0.784"]
M [pos="0.477,0.790"]
U1 [latent,pos="0.484,0.453"]
X [exposure,pos="0.301,0.788"]
Y [outcome,pos="0.699,0.778"]
IV -> X
M -> Y
U1 -> X
U1 -> Y
X -> M
}')

ggdag(dag7)
```




```
####simulation code############################
set.seed(12345)
n <- 300
u1 <- rnorm(n,5,1)
iv <- rnorm(n,1,5)
x <- .9*u1 + .5*iv + rnorm(n,2,.2)
m <- .2*x + rnorm(n,0,.5)
y <- .9*u1 - .5*m + rnorm(n,0,.5)

dfex7 <- data.frame(iv,x,m,y)
###############################################

####analysis code##############################
#unadjusted model
lm.u <- lm(y~x)
summary(lm.u)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.5210 -0.7625 -0.0100  0.6849  3.3294
##
## Coefficients:
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 3.837204   0.174241  22.022   <2e-16 ***
## x           0.001308   0.022499   0.058    0.954
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.03 on 298 degrees of freedom
## Multiple R-squared:  1.134e-05,  Adjusted R-squared:  -0.003344
## F-statistic: 0.003379 on 1 and 298 DF,  p-value: 0.9537
```

The unadjusted model yields an effect that is close to zero, which is biased. I will first present some code that uses the IV (using a SEM). I will contrast this code with the (more common) two-stage-least-squares results.

```
mymodel <- 'y ~ x
            x ~ iv
            x ~~ y'

fit1 <- sem(mymodel,data=dfex7)
summary(fit1)
```

```
## lavaan (0.6-1) converged normally after  35 iterations
##
##   Number of observations                           300
##
##   Estimator                                         ML
##   Model Fit Test Statistic                       0.000
##   Degrees of freedom                                 0
##
## Parameter Estimates:
##
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##   Standard Errors                             Standard
##
## Regressions:
##                    Estimate  Std.Err  z-value  P(>|z|)
##   y ~
##     x                -0.119    0.025   -4.744    0.000
##   x ~
##     iv                0.501    0.011   45.413    0.000
##
## Covariances:
##                    Estimate  Std.Err  z-value  P(>|z|)
##  .y ~~
##    .x                 0.843    0.079   10.627    0.000
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)
##    .y                 1.156    0.103   11.175    0.000
##    .x                 0.888    0.072   12.247    0.000
```

```
summary(ivreg(y~x | iv, data=dfex7))
```

```
##
## Call:
## ivreg(formula = y ~ x | iv, data = dfex7)
```

```
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.6686 -0.7446 -0.0312  0.6724  3.3769
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.71439    0.19380  24.326  < 2e-16 ***
## x           -0.11920    0.02521  -4.728 3.51e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.079 on 298 degrees of freedom
## Multiple R-Squared: -0.09626,    Adjusted R-squared: -0.09994
## Wald test: 22.35 on 1 and 298 DF,  p-value: 3.505e-06
```

The SEM and the two-stage-least-squares model are identical (up to the second digit), and both indicate a small, negative treatment effect.

```
mymodel <- 'y ~ a*m
            m ~ b*x
            x ~~ y
            te := a*b'

fit2 <- sem(mymodel,data=dfex7)
summary(fit2)
```

```
## lavaan (0.6-1) converged normally after  21 iterations
##
##   Number of observations                           300
##
##   Estimator                                         ML
##   Model Fit Test Statistic                       0.000
##   Degrees of freedom                                 0
##   Minimum Function Value               0.0000000000000
##
## Parameter Estimates:
##
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##   Standard Errors                             Standard
##
## Regressions:
##                    Estimate  Std.Err  z-value  P(>|z|)
##   y ~
##     m          (a)   -0.626    0.117   -5.356    0.000
##   m ~
##     x          (b)    0.225    0.011   21.254    0.000
##
## Covariances:
##                    Estimate  Std.Err  z-value  P(>|z|)
##  .y ~~
##     x                 0.994    0.251    3.965    0.000
##
## Variances:
```

```
##                    Estimate  Std.Err  z-value  P(>|z|)
##    .y                 1.104    0.104   10.596    0.000
##    .m                 0.235    0.019   12.247    0.000
##     x                 6.991    0.571   12.247    0.000
##
## Defined Parameters:
##                    Estimate  Std.Err  z-value  P(>|z|)
##     te               -0.141    0.027   -5.194    0.000
```

The causal effect that is estimated through the mechanism, is also small and negative, and quite similar to the one that we estimated with the IV model. Both models identify the same causal effect, but using different methods.

Exercise:

1.) Try to think of some examples in your own field, where you could use an IV or a mechanism. Try to convince the person next to you why the untestable assumptions of either model may be satisfied.