# 21-18

Nikhil Gopal

4/17/2022

Nikhil Gopal/Kerem Tuncer unis: nsg2127, kt2716

**21.18**

*Working through your own example*: Continuing the example from the final exercises of the earlier chapters, consider a causal problem that it would make sense to estimate using instrumental variables. Perform the instrumental variables estimate, compare to the estimated causal effect using direct regression on the treatment variable (in both cases including relevant pre-treatment predictors in your regressions), and discuss the different assumptions of these different approaches.

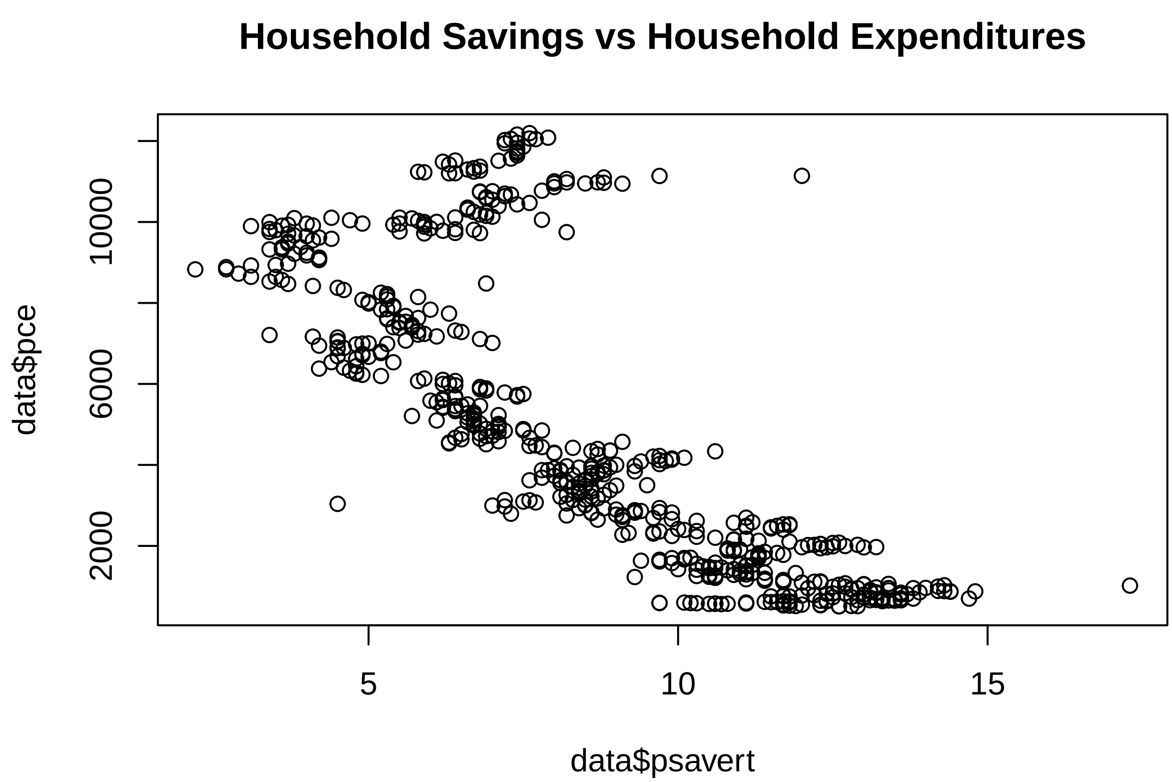
```
data <- read.csv("economics.csv")

data$unemploy_per_capita <- data$unemploy/data$pop
```

The above is an economics data set from the St. Louis Fed, which provides measurements of important economic indicators in the United States from 1967 to 2015. We are interested in being able to predict American unemployment levels. Two key indicators that would like affect unemployment rates are household savings (psavert) and personal consumption expenditure, which is a measure of household spending and an indirect proxy measurement of the cost of living. This dataset gives unemployment measured by number of unemployed Americans. To adjust for the increase in US population over time, we created a new column that represents unemployment per capita.

This is the "economics" dataset that is built into R. We were having trouble loading it in Markdown, so we simply wrote the dataset as a csv to our computers and then continued.

The hypothesis is that higher cost of living would be a predictor of unemployment, as rising household expenditure would motivate people to work who otherwise would not. However, savings rate also would affect household expenditure, as a larger proportion of people who have high levels of savings might choose to not work and live off their interest, which directly affects unemployment but slightly decreases household savings. These variables are also extremely correlated (correlation = -0.79), which makes them suitable candidates for instrument variable analysis:

```r
plot(data$psavert, data$pce, main = "Household Savings vs Household Expenditures")
```




```r
cor(data$psavert, data$pce)
```

```
## [1] -0.7928546
```

Below we will fit the models and perform the analysis:

treatment = pce instrument = savings outcome = unemployment

```r
fit_instrument <- stan_glm(psavert ~ pce, data = data, refresh = 0)
data$household_savings_hat <- fit_instrument$fitted

fit_outcome <- stan_glm(unemploy_per_capita ~ household_savings_hat, data = data, refresh = 0)

print(fit_outcome)
```

```
## stan_glm
##  family:       gaussian [identity]
##  formula:      unemploy_per_capita ~ household_savings_hat
##  observations: 574
##  predictors:   2
## ------
##                       Median MAD_SD
```

```
## (Intercept)           0.0    0.0
## household_savings_hat 0.0    0.0
##
## Auxiliary parameter(s):
##       Median MAD_SD
## sigma 0.0    0.0
##
## ------
## * For help interpreting the printed output see ?print.stanreg
## * For info on the priors used see ?prior_summary.stanreg
```

It appears this model does not fit well, it gave a coefficient estimate of zero. Let us try adding a covariate, this time the duration of unemployment variable:

```
fit_instrument <- stan_glm(psavert ~ pce + uempmed, data = data, refresh = 0)
data$household_savings_hat <- fit_instrument$fitted

fit_outcome <- stan_glm(unemploy_per_capita ~ household_savings_hat + uempmed, data = data, refresh = 0

print(fit_outcome)
```

```
## stan_glm
##  family:       gaussian [identity]
##  formula:      unemploy_per_capita ~ household_savings_hat + uempmed
##  observations: 574
##  predictors:   3
## ------
##                       Median MAD_SD
## (Intercept)           0.0    0.0
## household_savings_hat 0.0    0.0
## uempmed               0.0    0.0
##
## Auxiliary parameter(s):
##       Median MAD_SD
## sigma 0.0    0.0
##
## ------
## * For help interpreting the printed output see ?print.stanreg
## * For info on the priors used see ?prior_summary.stanreg
```

Finally, lets try fitting a model without an instrumental variable:

```
fit <- stan_glm(unemploy_per_capita ~ pce, data = data, refresh = 0)

print(fit)
```

```
## stan_glm
##  family:       gaussian [identity]
##  formula:      unemploy_per_capita ~ pce
##  observations: 574
##  predictors:   2
## ------
```

```
##             Median MAD_SD
## (Intercept) 0.0    0.0
## pce         0.0    0.0
##
## Auxiliary parameter(s):
##       Median MAD_SD
## sigma 0.0    0.0
##
## ------
## * For help interpreting the printed output see ?print.stanreg
## * For info on the priors used see ?prior_summary.stanreg
```

All of the models had coefficient estimates of zero. It appears that neither an instrumental nor a traditional regression showed a relationship between PCE and US Unemployment per capita. The instrumental model assumes that household savings predicts household expenditures, and that household savings is randomly assigned, allowing the researcher to establish causality between the treatment of interest and the outcome variable.

While there appears to be no association between the variables anyways, it may have been a faulty assumption to assume that household savings was randomly assigned. Initially we thought it would be random, but many variables affect household savings, so an instrumental model was probably not appropriate for this analysis.