

Practical: Instrumental Variables

Qingyuan Zhao (Statistical Laboratory)*

July 25, 2025

This practice requires the `ivmodel` package:

```
# install.packages("ivmodel") # execute if ivmodel is not already installed.
library(ivmodel)
```

Instrumental variables are commonly used to deal with unmeasured confounding. We demonstrate this method using a prominent study on the economic return of schooling (Card, D. (1993), *Using Geographic Variation in College Proximity to Estimate the Return to Schooling*, Technical Report w4483, National Bureau of Economic Research, Cambridge, MA.).

1. Read the abstract of Card (1993). Familiarize yourself with this dataset (you can find the relevant information by executing `help(card.data)`) and the `ivmodel` function (`help(ivmodel)`).
2. (Group discussion) Execute the following code in R.

```
data(card.data)
# One instrument #
Y=card.data[, "lwage"]
D=card.data[, "educ"]
Z=card.data[, "nearc4"]
Xname=c("exper", "expersq", "black", "south", "smsa", "reg661",
        "reg662", "reg663", "reg664", "reg665", "reg666", "reg667",
        "reg668", "smsa66")
X=card.data[, Xname]
card.model1IV = ivmodel(Y=Y, D=D, Z=Z, X=X)
summary(card.model1IV)
```

Which variables in the dataset are used as the exposure, outcome, and instrumental variable? What covariates are used in this IV model? Discuss the output from the code.

3. (Group discussion) Which IV model does the following code uses? Compare the results of these two models.

```
Z = card.data[, c("nearc4", "nearc2")]
card.model2IV = ivmodel(Y=Y, D=D, Z=Z, X=X)
card.model2IV
```

*qyzhao@statslab.cam.ac.uk