14.74 Recitation 5

Partial Compliance, Instrumental Variables

Maheshwor Shrestha

October 11, 2013

Agenda

- Randomized experiments with partial compliance
- Instrumental Variables
  - Wald Estimates
  - Indirect Least Squares (ILS)
  - Two-stage Least Squares (2SLS)

Randomized Experiments with Partial Compliance

- Deworming experiment in Kenya
  - Schools randomized to receive deworming pills
  - only children who showed up for schools on that day took the pill
  - You want to study the effect of deworming pill on school attendance
- CARES program offered accounts to help people quit smoking
  - On a sample of smokers, some individuals were randomly chosen to get an offer to open such accounts
  - Only a fraction of them actually opened the account
  - You want to study the effect of the account on smoking rate
- Education voucher programs
  - Vouchers subsidizing schooling provided randomly to school children
  - Not all children who got the voucher went to school
  - You want to study the effect of schooling on wages (when children grow up)

Wrong Estimates

- Suppose you compared the outcomes of those who took up the program with those who didn't
  - Schooling attendance of children who took the pill with those who didn't
  - Smoking rates of those who opened the account with those who didn't
  - Wages of those who used the voucher with those who didn't

What is wrong with these comparisons?

Notation

- Let $Z_i \in \{0, 1\}$ represent whether an individual was offered the treatment or not
  - $Z_i = 1$ refers to the treatment group and $Z_i = 0$ as the control
  - In the examples above, Z is randomly assigned
- Let $W_i \in \{0, 1\}$ represent whether the individual took up the treatment or not.
  - But W_i is an outcome that is affected by whether they were offered the treatment or not
  - Let $W_i(z)$ represent the potential treatment status when $Z_i = z$
    - So, for the treatment group $W_i(1)$ is observed and $W_i(0)$is a counterfactual and for the control group $W_i(0)$ is observed and $W_i(1)$ is a counterfactual

Notation (cont...)

- Let Y_i represent the realized outcomes
  - But Y_i may be affected by whether the individuals were in treatment group or not and by whether they took up the treatment or not
  - Write $Y_i(w, z)$ as the potential outcome of an individual in treatment group z and with treatment-status w
  - What is observed and what is not?
- This is a more general setting where outcome is affected by what is offered as well as the treatment status

Problematic comparison

- What was the problem of comparing means between the treated (in the treatment group) and untreated (in the control group)?

$$
\begin{aligned}
& E\left[Y_i | W_i = 1, Z_i = 1\right] - E\left[Y_i | W_i = 0, Z_i = 0\right] \\
= & E\left[Y_i(1,1) \,|\, W_i(1) = 1, Z_i = 1\right] - E\left[Y_i(0,0) \,|\, W_i(0) = 0, Z_i = 0\right] \\
= & E\left[Y_i(1,1) \,|\, W_i(1) = 1\right] - E\left[Y_i(0,0) \,|\, W_i(0) = 0\right] \\
= & E\left[Y_i(1,1) \,|\, W_i(1) = 1\right] - E\left[Y_i(0,0) \,|\, W_i(1) = 1\right] \\
& + E\left[Y_i(0,0) \,|\, W_i(1) = 1\right] - E\left[Y_i(0,0) \,|\, W_i(0) = 0\right]
\end{aligned}
$$

  - The second term is the selection bias in the comparison.
  - In the examples, why could there be a selection bias?
    - What type of people would choose to takeup the treatment (and have $W_i = 1$)?
    - What kind of bias does that lead to?

Intent to Treat

- Now compare the means by completely ignoring the takeup, and just looking at the outcomes by offer status (ie, Z_i)

$$\begin{aligned} E\left[Y_i|Z_i=1\right]-E\left[Y_i|Z_i=0\right] &= E\left[Y_i\left(\cdot,1\right)|Z_i=1\right]-E\left[Y\left(\cdot,0\right)|Z_i=0\right] \\ &= E\left[Y_i\left(\cdot,1\right)|Z_i=1\right]-E\left[Y\left(\cdot,0\right)|Z_i=1\right] \\ &\quad + E\left[Y\left(\cdot,0\right)|Z_i=1\right]-E\left[Y\left(\cdot,0\right)|Z_i=0\right] \end{aligned}$$

  - What is the selection bias? How does randomization help us?
  - Since randomization ensures that offer is independent of all potential outcomes, $E\left[Y\left(\cdot,0\right)|Z_i=1\right]-E\left[Y\left(\cdot,0\right)|Z_i=0\right]=0$
- Comparison of the means by offer status is called the **intent to treat** effect (ITT).
  - This is the difference in mean by whom we intended to treat
  - We just saw that if Z_i is random, ITT is causal.

Intent to Treat (cont...)

- How do you get the reduced form estimate by a regression?

$$y_i = \alpha + \beta Z_i + \varepsilon_i$$

- β gives us the ITT estimate:

$$E\left[Y_i | Z_i = 1\right] - E\left[Y_i | Z_i = 0\right]$$

- Is this comparison something we are interested in? Who might be interested in this estimate
- Is this what we want?
  - Ideally, we would like to know the effect of the treatment on those who actually took the program. In our examples, who are they?
  - ITT is, in some sense, diluted by all those people who did not get the treatment

Takeup rates

- We can do another comparison

$$E\left[W_i|Z_i=1\right]-E\left[W_i|Z_i=0\right]$$

  this is the difference in takeup rates between the treatment and control group.
  - What is $E\left[W_i|Z_i=0\right]$ in our examples?
- This is called the **first stage**. In case of RCTs, it is the difference in takeup-rate
  - Is this causal?
    - Yes, similar proof as with the ITT (thanks to randomization)
- Why might this be interesting?
- Is this what we want?

Wald Estimate

- Since ITT is diluted by all individuals not taking up the program when offered (or individuals who are not offered the program also somehow are treated), we want to scale it up by the increase in takeup between two. This is the **Wald Estimate**

$$\hat{\beta}_{WALD} = \frac{E\left[Y_i|Z_i=1\right] - E\left[Y_i|Z_i=0\right]}{E\left[W_i|Z_i=1\right] - E\left[W_i|Z_i=0\right]} = \frac{ITT}{\Delta \textbf{Takeup rate}}$$

- Under what assumption is the Wald estimate saying anything causal?

Assumptions

- Under additional assumptions
  1. **First stage**: $E\left[W_i|Z_i=1\right]-E\left[W_i|Z_i=0\right]\neq 0$
  2. **Exclusion**: The offer Z_i affects the outcome only through actual treatment status W_i. That is, there is not direct effect of the offer on the outcome

$$Y_i\left(1,0\right)=Y_i\left(1,1\right)\text{ and }Y_i\left(0,1\right)=Y_i\left(0,0\right)$$

  the wald estimate is same as the instrumental variable estimate
    - This is a strong assumption
    - In our examples, what would violate the assumptions?

Interpretation of Wald

- Under the assumptions of monotonicity and exclusion

$$\hat{\beta}_{WALD} = \frac{E\left[Y_i|Z_i = 1\right] - E\left[Y_i|Z_i = 0\right]}{E\left[W_i|Z_i = 1\right] - E\left[W_i|Z_i = 0\right]}$$

gives the average treatment effect on the treated

Motivation for Instruments

- Suppose we are interested in causally estimating β in the following regression
$$y_i = \alpha + \beta W_i + \gamma X_i + \varepsilon_i$$
To set our minds, think of y_i as wages and W_i as number of years of schooling (not binary as before) X_i as other controls (experience, for example)
- Suppose we ran the regression using cross-section data. Will we get a causal estimate for β? Why or why not?
  - Omitted Variable Bias - ability, institutions, etc
  - Reverse causality
  - Measurement error

Motivation

- What does this mean in terms of OLS assumptions?
  - $Cov\left(W_i, \varepsilon_i\right) \neq 0$
  - That is W_i is **endogenous**
- So, to estimate β causally we need something that changes education without directly affecting any of the other channels (ability, institution, income, measurement errors)
  - That is, we need something (an instrument) that exogenously changes education
  - We will call this instrument Z_i

Assumptions

- There must be a **first-stage**

$$W_i = \pi_0 + \pi_1 Z_i + \gamma X_i + \eta_i$$

with $\pi_1 \neq 0$

  - The instrument must affect the endogenous variable
  - The instrument is exogenous ($Cov\left(Z_i, \eta_i\right) = 0$)
    - no OVB, reverse causality etc

- The **exclusion restriction** must be satisfied
  - instrument must not have a direct effect on the outcome, except through the channel of W_i
  - That is: $Cov\left(Z_i, \varepsilon_i\right) = 0$

Wald Estimate

- Under these assumptions and in the case the instrument is binary (note that the endogenous variable W_i need not be binary), we have (ignoring controls)

$$
\begin{aligned}
y_i =& \alpha + \beta W_i + \varepsilon_i \\
E\left[y_i | Z_i = 1\right] =& \alpha + \beta \cdot E\left[W_i | Z_i = 1\right] + E\left[\varepsilon_i | Z_i = 1\right] \\
E\left[y_i | Z_i = 0\right] =& \alpha + \beta \cdot E\left[W_i | Z_i = 0\right] + E\left[\varepsilon_i | Z_i = 0\right]
\end{aligned}
$$

  Subtracting

$$
\begin{aligned}
E\left[y_i | Z_i = 1\right] - E\left[y_i | Z_i = 0\right] =& \beta \left(E\left[W_i | Z_i = 1\right] - E\left[W_i | Z_i = 0\right]\right) \\
& + \left(E\left[\varepsilon_i | Z_i = 1\right] - E\left[\varepsilon_i | Z_i = 0\right]\right) \\
\hat{\beta}_W =& \frac{E\left[y_i | Z_i = 1\right] - E\left[y_i | Z_i = 0\right]}{E\left[W_i | Z_i = 1\right] - E\left[W_i | Z_i = 0\right]}
\end{aligned}
$$

- Where do the assumptions come in?

Wald Estimate (cont...)

$$\hat{\beta}_W = \frac{E\left[y_i|Z_i = 1\right] - E\left[y_i|Z_i = 0\right]}{E\left[W_i|Z_i = 1\right] - E\left[W_i|Z_i = 0\right]}$$

- The numerator is called the **reduced form**
  - In the RCT language, this is the ITT estimate
- The denominator is call the **first-stage**
  - In the RCT language, this is the difference in takeup

Indirect Least Squares (ILS)

- We have the two equations

$$\begin{aligned} y_i =& \alpha + \beta W_i + \varepsilon_i \\ W_i =& \pi_0 + \pi_1 Z_i + \eta_i \end{aligned}$$

- Plugging W_i in the first equation

$$\begin{aligned} y_i =& \alpha + \beta \left(\pi_0 + \pi_1 Z_i + \eta_i\right) + \varepsilon_i \\ =& \alpha + \beta\pi_0 + \beta\pi_1 Z_i + \beta\eta_i + \varepsilon_i \\ =& \theta_0 + \theta_1 Z_i + \xi_i \end{aligned}$$

- What assumptions are necessary to estimate the relationship $y_i = \theta_0 + \theta_1 Z_i + \xi_i$?
  - How do the assumptions help

ILS

- The assumption $Cov\left(Z_i, \xi_i\right) = 0$ requires that the instrument is exogenous $Cov\left(Z_i, \eta_i\right) = 0$ and the exclusion restriction holds $Cov\left(Z_i, \varepsilon_i\right) = 0$
- It is clear that $\theta_1 = \beta\pi_1$. Hence

$$\hat{\beta}_I = \frac{\hat{\theta}_1}{\hat{\pi}_1} = \frac{\frac{Cov(y_i,Z_i)}{Var(Z_i)}}{\frac{Cov(W_i,Z_i)}{Var(Z_i)}} = \frac{Cov\left(y_i, Z_i\right)}{Cov\left(W_i, Z_i\right)}$$

2-Stage Least Squares

- We have the two equations

$$\begin{aligned} y_i =& \alpha + \beta W_i + \varepsilon_i \\ W_i =& \pi_0 + \pi_1 Z_i + \eta_i \end{aligned}$$

- First estimate

$$W_i = \pi_0 + \pi_1 Z_i + \eta_i$$

  and get the predicted value $\hat{W}_i = \hat{\pi}_0 + \hat{\pi}_1 Z_i$
- Next estimate

$$y_i = \alpha + \beta \hat{W}_i + \varepsilon_i$$

  when will this equation be identified?

2SLS

- Identification requires that $Cov\left(\hat{W}_i, \varepsilon_i\right) = 0$ or $Cov\left(\hat{\pi}_0 + \hat{\pi}_1 Z_i, \varepsilon_i\right) = 0$. Which assumption helps here?
- Why did we need the other assumption?
- When there is one instrument, both methods (all three if instrument is binary) give the exact same results
- When there are multiple instruments, 2SLS gives optimally weighted average of all possible linear combinations of various estimator
- So, ALWAYS do 2SLS
    - In R use the commad: **ivreg** in the package AER
- NEVER do 2SLS *manually* in two-stages. Why?
    - The second stage standard errors will be wrong as we did not take into account that $\hat{W}_i$ is estimated