Quant II

Lab 4: IV and Inference

Giacomo Lemoli

February 28, 2022

Today's plan

- HW2 review
- Inference and standard errors
- Instrumental Variables

Housekeeping

- Deadlines
- Data for HW3

Homework 2

Preamble: plug-in estimation

A general principle for estimating theoretical properties of distributions, e.g. moments (cf. Aronow and Miller (2019), sec.3.3)

Plug-in principle: We want to estimate some parameter $\theta = T(F)$, where F is a CDF and $T : \mathcal{F} \to \mathbb{R}$ is a functional mapping CDFs into real numbers. A plug-in estimator for θ is $\hat{\theta} = T(\hat{F})$, where $\hat{F}$ is an empirical CDF.

Under some regularity conditions of T, $\hat{\theta} = T(\hat{F})$ has desirable properties such as asymptotic normality.

Basics of inference: a friendly review

- We estimate some theoretically interesting quantity, we want to know how far we may be from the true value
- Population quantity estimated from random sample (e.g. E(X), PATE) or sample quantity estimated from random assignment (e.g. SATE)
- "Uncertainty" of our estimate depends on the sampling distribution of our estimator, in particular on its sampling standard deviation (the standard error), which we don't know
- For instance, if we know that $X \sim N(\mu, \sigma^2)$, then we know that $\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$ always
- If we don't know the actual distribution of our variables, we usually can't rely on analytical derivations of the sampling distribution

Basics of inference: a friendly review

- We estimate some theoretically interesting quantity, we want to know how far we may be from the true value
- Population quantity estimated from random sample (e.g. E(X), PATE) or sample quantity estimated from random assignment (e.g. SATE)
- "Uncertainty" of our estimate depends on the sampling distribution of our estimator, in particular on its sampling standard deviation (the standard error), which we don't know
- For instance, if we know that $X \sim N(\mu, \sigma^2)$, then we know that $\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$ always
- If we don't know the actual distribution of our variables, we usually can't rely on analytical derivations of the sampling distribution
  - Use asymptotic approximations using familiar results
  - Impose distributional assumptions

Basics of inference: a friendly review

Aronow and Miller (2019), p.125

A valid confidence interval for a parameter θ with coverage $(1-\alpha)$ is a random interval $CI_{(1-\alpha)}(\theta)$ such that $P[\theta \in CI_{(1-\alpha)}(\theta)] \geq 1-\alpha$

An asymptotically valid confidence interval is one such that $\lim_{n \to -\infty} P[\theta \in CI_{(1-\alpha)}(\theta)] \geq 1-\alpha$

If our estimator $\hat{\theta}$ has an asymptotically normal distribution and $\hat{V}(\hat{\theta})$ is consistent for $V(\hat{\theta})$, then an asymptotically valid confidence interval for $\hat{\theta}$ is:

$$CI_{(1-\alpha)}(\theta) = [\hat{\theta} - z_{(1-\frac{\alpha}{2})}\sqrt{\hat{V}(\hat{\theta})}, \hat{\theta} + z_{(1-\frac{\alpha}{2})}\sqrt{\hat{V}(\hat{\theta})}]$$

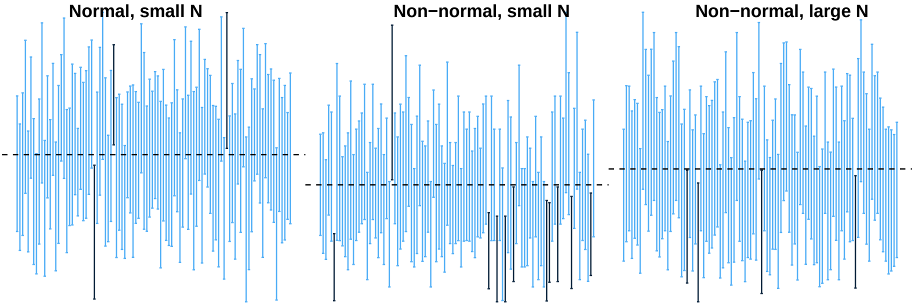
An application of the plug-in principle: if $\hat{V}(\hat{\theta}) \to V(\theta)$, $Z' = \left(\frac{\hat{\theta}-\theta}{\sqrt{\hat{V}(\hat{\theta})}}\right) \to Z = \left(\frac{\hat{\theta}-\theta}{\sqrt{V(\hat{\theta})}}\right) \sim N(0,1)$.

Example: normal approximations for confidence intervals

```r
library(ggplot2)
set.seed(123)

# Number of confidence intervals
tot <- 100

s1 <- s2 <- s3 <- matrix(NA, tot, 2)

confint <- function(x, alpha){
  est <- mean(x)
  lower <- est - qnorm(1-alpha/2)*(sd(x)/sqrt(length(x)))
  upper <- est + qnorm(1-alpha/2)*(sd(x)/sqrt(length(x)))
  c(lower, upper)
}

for (i in 1:tot){
  s1[i,] <- confint(rnorm(30, mean = 0.5), 0.05)
  s2[i,] <- confint(rpois(30, 0.5), 0.05)
  s3[i,] <- confint(rpois(100, 0.5), 0.05)
}

plot <- function(m){
  require(dplyr)
  plot <- m %>% as.data.frame() %>% mutate(id = c(1:dim(m)[1])) %>%
    ggplot(aes(x=id, col=ifelse(V1<=0.5 & V2>=0.5, 1, 0))) + geom_errorbar(aes(ymin = V1, ymax = V2)) +
    geom_hline(yintercept = 0.5, linetype = "dashed") + theme_void() + theme(legend.position = "none")
  plot
}
```

Example: normal approximation for confidence intervals

```
library(ggpubr)
ggarrange(plot(s1), plot(s2), plot(s3), ncol = 3,
          labels = list("Normal, small N", "Non-normal, small N", "Non-normal, large N"))
```



Standard errors for causal effects

Imagine a randomized experiment where the PATE is our target. Our estimator is the difference in means $\hat{\tau} = \bar{Y}_1 - \bar{Y}_0$.

We have seen that the Neyman estimator is an unbiased estimator for the variance of $\hat{\tau}$ under sampling and randomization distribution:

$$\hat{V}_{ney} = \frac{\hat{s}^2_{Y_1}}{n_1} + \frac{\hat{s}^2_{Y_0}}{n_0}$$

Therefore we can use the Neyman standard error $\sqrt{\hat{V}_{ney}}$ to construct confidence intervals around $\hat{\tau}$ and test hypotheses about $\hat{\tau}_{PATE}$

Equivalence of Neyman SE and robust regression SE

The formulation of the Neyman estimator provides an intuitive rational for heteroskedasticity-robust SEs: it allows outcome variance to vary with levels of treatment (the definition of heteroskedasticity).

Indeed, the "HC2" estimator for robust regression SEs is equivalent to it.

To see this, let's use `estimatr`, a package for the analysis of randomized experiments, where HC2 is the default.

Equivalence of Neyman SE and robust regression SE

```r
## Simulate a randomized experiment
set.seed(123)
library(estimatr)

# Simulated population
pop <- data.frame(Y1 = rnorm(1000, 4, 2), Y0 = rnorm(1000, 0.5, 3))

# Random sample
sample <- pop[sample(nrow(pop), 100),]
sample$D <- 0
sample$D[sample(100, 30)] <- 1

# Observed potential outcomes
sample <- sample %>% mutate(Y = D*Y1 + (1-D)*Y0)

# Compute the Neyman SE manually
(ney_se <- sqrt(var((sample$Y[sample$D==1]))/sum(sample$D==1) + var((sample$Y[sample$D==0]))/sum(sample$D==0)))
```

```
## [1] 0.513596
```

```r
# Regression with HC2 robust SE
lm_robust(Y ~ D, data = sample)
```

```
##              Estimate Std. Error   t value     Pr(>|t|)   CI Lower CI Upper DF
## (Intercept) 0.3137636  0.3844703 0.8160932 4.164261e-01 -0.4492052 1.076732 98
## D           4.1641379  0.5135960 8.1078083 1.493054e-12  3.1449234 5.183352 98
```

```r
# Note this is different from HC1 (the Stata default)
lm_robust(Y ~ D, data = sample, se_type = "stata")
```

```
##              Estimate Std. Error   t value     Pr(>|t|)   CI Lower CI Upper DF
## (Intercept) 0.3137636  0.3855896 0.8137243 4.177757e-01 -0.4514264 1.078954 98
## D           4.1641379  0.5128987 8.1188318 1.414176e-12  3.1463072 5.181969 98
```

Robust variance estimators for the Difference in Means

See Imbens and Rubin Ch.7 and MHE Ch.8 (be aware of notation)

$\hat{s}^2_{Y_j} = \frac{\sum_{i:D_i=j}^{N_j}(Y_i - \bar{Y}_j)^2}{n_j - 1}$

- Non-robust/conventional: $\hat{s}^2_Y(\frac{1}{n_1} + \frac{1}{n_0})$
- HC0: $\frac{n_0-1}{n_0^2}\hat{s}^2_{Y_0} + \frac{n_1-1}{n_1^2}\hat{s}^2_{Y_1}$
- HC1: $\frac{n}{n-2}\left(\frac{n_0-1}{n_0^2}\hat{s}^2_{Y_0} + \frac{n_1-1}{n_1^2}\hat{s}^2_{Y_1}\right)$
- HC2: $\frac{\hat{s}^2_{Y_1}}{n_1} + \frac{\hat{s}^2_{Y_0}}{n_0}$
- HC3: $\frac{\hat{s}^2_{Y_1}}{n_1-1} + \frac{\hat{s}^2_{Y_0}}{n_0-1}$

Robust variance estimators in regression

In multivariate regression, the problem generalizes to the estimation of a covariance *matrix* for the vector of OLS coefficients.

The general form of a sampling variance estimator for the OLS vector $\widehat{\beta}$ is

$$\hat{V}(\hat{\beta}) = (X'X)^{-1}X'\hat{\Psi}X(X'X)^{-1}$$

where $\hat{\Psi}$ is an estimator of $plim[\epsilon\epsilon']$. Estimators differ in what goes in the diagonal of $\hat{\Psi}$.

Robust variance estimators in regression

From MHE, Ch.8.

- Non-robust/conventional: $\hat{\psi}_i = \hat{\psi} = \frac{\sum_{i=1}^{n} e_i^2}{n-k} = \hat{\sigma}$
- HC0: $\hat{\psi}_i = \hat{e}_i^2$
- HC1: $\hat{\psi}_i = \frac{n}{n-k} \hat{e_i}^2$
- HC2: $\hat{\psi} = \frac{1}{1-h_{ii}} \hat{e}_i^2$
- HC3: $\hat{\psi} = \frac{1}{(1-h_{ii})^2} \hat{e}_i^2$

Robust standard errors in R

```r
# lm + lmtest + sandwich
library(sandwich); library(lmtest)
fit <- lm(Y ~ D, data = sample)
coeftest(fit, vcov = vcovHC(fit, type = "HC2"))
```

```
##
## t test of coefficients:
##
##             Estimate Std. Error t value  Pr(>|t|)
## (Intercept)  0.31376    0.38447  0.8161    0.4164
## D            4.16414    0.51360  8.1078 1.493e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Robust standard errors in R

```r
# Manual (only the conventional case)
summary(fit)
```

```
##
## Call:
## lm(formula = Y ~ D, data = sample)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.7960 -2.1130 -0.0758  2.4649  6.0472
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.3138     0.3446   0.910    0.365
## D             4.1641     0.6292   6.618 1.95e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.884 on 98 degrees of freedom
## Multiple R-squared:  0.3089, Adjusted R-squared:  0.3018
## F-statistic: 43.79 on 1 and 98 DF,  p-value: 1.949e-09
```

```r
N <- nrow(sample)
D <- cbind(rep(1, N), sample$D)
K <- dim(D)[2]

vcov <- solve((t(D) %*% D)) %*% t(D) %*% diag((sum(residuals(fit)^2)/(N-K)), N, N) %*% D %*% solve((t(D) %*% D)

cbind(coef(fit), sqrt(diag(vcov)))
```

```
##                  [,1]      [,2]
## (Intercept) 0.3137636 0.3446480
## D           4.1641379 0.6292383
```

Robust standard errors in R and Stata

- Some packages like `estimatr` or `fixest` have built-in options for various types of robust SEs
- Like `lm`, other packages only provide the default homoskedastic SEs, so you have to call a post-estimation command that computes the robust covariance matrix
- Looks annoying if you come from Stata
- **always** check the `help` material to see what SEs are computed
- In Stata: `regress y x, vce()`, the argument of `vce()` can be `robust`, `hc2`, `hc3`, `bootstrat`, `cluster clustervar`
- If you replicate analyses in another language, cross-check the documentation, as different adjustments may be applied (see below for IV)

Bootstrap standard errors in R and Stata

- In R: check the `boot` package and `car::Boot()`. More in this article
- If you are skilled coders you can write your own bootstrap function for regression (just be very careful)
- Cluster bootstrap SEs: check R's `fwildclusterboot` and Stata's `boottest`
  - Note they don't produce confidence intervals, only the p-values
- **Important**: since bootstrap randomizes, set seed or check what your package is doing

Instrumental Variables

Principal Strata:

- Compliers: $D(1) = 1, D(0) = 0$
- Always-takers $D(1) = D(0) = 1$
- Never-takers $D(1) = D(0) = 0$
- Defiers: $D(1) = 0, D(0) = 1$

IV in practice: Dower et al (2018)

American Political Science Review (2018) 112, 1, 125–147
doi:10.1017/S0003055417000454

Collective Action and Representation in Autocracies: Evidence from Russia's Great Reforms

PAUL CASTAÑEDA DOWER *University of Wisconsin–Madison*
EVGENY FINKEL *George Washington University*
SCOTT GEHLBACH *University of Wisconsin–Madison*
STEVEN NAFZIGER *Williams College*

We explore the relationship between capacity for collective action and representation in autocracies with data from Imperial Russia. Our primary empirical exercise relates peasant representation in new institutions of local self-government to the frequency of peasant unrest in the decade prior to reform. To correct for measurement error in the unrest data and other sources of endogeneity, we exploit idiosyncratic variation in two determinants of peasant unrest: the historical incidence of serfdom and religious polarization. We find that peasants were granted less representation in districts with more frequent unrest in preceding years—a relationship consistent with the Acemoglu-Robinson model of political transitions and inconsistent with numerous other theories of institutional change. At the same time, we observe patterns of redistribution in subsequent years that are inconsistent with the commitment mechanism central to the Acemoglu-Robinson model. Building on these results, we discuss possible directions for future theoretical work.

The effect of peasant unrest on representation

```
library(haven); library(AER); library(stargazer)
data <- read_dta("DFGN_cleaned.dta")

## OLS
olsfit <- lm(peasantrepresentation_1864 ~ afreq + distance_moscow + goodsoil + lnurban + lnpopn + province_capi

## IV (1): serfdom
ivfit1 <- ivreg(peasantrepresentation_1864 ~ afreq + distance_moscow +
                goodsoil + lnurban + lnpopn + province_capital | serfperc1 +
                distance_moscow + goodsoil + lnurban + lnpopn +
                province_capital, data=data)

## IV (2): religious polarization
ivfit2 <- ivreg(peasantrepresentation_1864 ~ afreq + distance_moscow +
                goodsoil + lnurban + lnpopn + province_capital | religpolarf4_1870 +
                distance_moscow + goodsoil + lnurban + lnpopn + province_capital, data=data)

mod <- list(olsfit, ivfit1, ivfit2)
ses <- lapply(mod, function(x) coeftest(x, vcov = vcovHC(x, type = "HC1"))[,"Std. Error"])
labs <- c("", "Z: % serfs", "Z: religious pol.")
```

The effect of peasant unrest on representation

```
stargazer(mod, se = ses, column.labels = labs, omit.stat = c("f", "ser"), type = "text")
```

```
##
## =========================================================
##                           Dependent variable:
##                  ----------------------------------------
##                         peasantrepresentation_1864
##                     OLS            instrumental
##                                      variable
##                         Z: % serfs Z: religious pol.
##                     (1)       (2)            (3)
## ---------------------------------------------------------
## afreq            -4.249**  -41.999***     -32.770*
##                   (1.830)   (8.509)       (17.352)
##
## distance_moscow    0.379   -7.222***       -5.401
##                   (1.288)   (2.203)        (3.733)
##
## goodsoil           1.127    3.860***       3.101*
##                   (0.811)   (1.317)        (1.801)
##
## lnurban          -2.605*** -1.901***      -2.086***
##                   (0.439)   (0.584)        (0.555)
##
## lnpopn           5.224***   8.291***      7.597***
##                   (1.092)   (1.243)        (1.777)
##
## province_capital -3.345*** -5.177***      -4.689***
##                   (1.281)   (1.679)        (1.642)
##
## Constant           3.633    -23.952*       -17.715
##                  (12.079)   (13.245)      (16.979)
##
```

Characterize the compliers

There are different strategies: see the discussion in Marbach and Hangartner (2020).

Let's go through a regression-based alternative to k-weights, also based on Abadie.

Call X a variable for which we want to know the mean among the treated compliers. D is the endogenous treatment and Z the instrument. Create $X^* = X * D$ and use it as outcome in the 2SLS model. The coefficient of D instrumented is the mean of Z among the treated compliers.

Use $X * (1 - D)$ and replace D with $1 - D$ in the regression to do the same for the untreated compliers.

Characterize the unrest compliers

```r
# Variables used in the analysis
vars <- c("mliteracy_1897", "lingfrac_1897", "popdens1858",
          "percnobleown_1877", "districtfemale_1863")
exvars <- c("distance_moscow", "goodsoil", "lnurban",
            "lnpopn", "province_capital")

# For simplicity, we collapse D and Z to a binary variable
data$D <- ifelse(data$afreq >= quantile(data$afreq, probs = 0.75, na.rm = T), 1, 0)
data$Z1 <- ifelse(data$serfperc1 >= quantile(data$serfperc1, 0.75, na.rm=T), 1, 0)
data$Z2 <- ifelse(data$religpolarf4_1870 >= quantile(data$religpolarf4_1870, 0.75, na.rm=T), 1, 0)
```

Characterize the unrest compliers

```r
##  Compliers' distribution, regression method
mean_comp <- function(var, Z){
  require(AER); require(dplyr)
  data <- mutate(data, Xc = get(var)*D)
  formula <- paste0("Xc", "~", "D", "+", paste(exvars, collapse = " + "),
                    "|", Z, "+", paste(exvars, collapse = " + "))
  fit <- ivreg(as.formula(formula), data=data)
  return(coef(fit)["D"])
}

means_z1_t <- means_z2_t <- rep(NA, length(vars))

# Population distribution
means_full <- round(apply(data[data$zemstvo==1,vars], 2, function(x) mean(x, na.rm=T)), 3)

# Compliers of % serf
for(i in 1:length(vars)){means_z1_t[i] <- round(mean_comp(vars[i], "Z1"), 3)}

# Compliers of rel. pol.
for(i in 1:length(vars)){means_z2_t[i] <- round(mean_comp(vars[i], "Z2"), 3)}
```

Characterize the unrest compliers

```
cbind(means_full, means_z1_t, means_z2_t)
```

```
##                      means_full means_z1_t means_z2_t
## mliteracy_1897           47.046     44.819     92.626
## lingfrac_1897             0.183     -0.405      1.794
## popdens1858              68.511     34.258    139.561
## percnobleown_1877        27.389     76.015     29.867
## districtfemale_1863   58729.526  31064.884  10005.480
```

Another example: Bautista et al (2020)

AMERICAN JOURNAL of POLITICAL SCIENCE

The Geography of Repression and Opposition to Autocracy

Maria Angélica Bautista University of Chicago
Felipe González Pontificia Universidad Católica de Chile
Luis R. Martínez University of Chicago
Pablo Muñoz FGV EPGE Brazilian School of Economics and Finance
Mounu Prem Universidad del Rosario

Abstract: *State repression is a prominent feature of nondemocracies, but its effectiveness in quieting dissent and fostering regime survival remains unclear. We exploit the location of military bases before the coup that brought Augusto Pinochet to power in Chile in 1973, which is uncorrelated to precoup electoral outcomes, and show that counties near these bases experienced more killings and forced disappearances at the hands of the government during the dictatorship. Our main result is that residents of counties close to military bases both registered to vote and voted "No" to Pinochet's continuation in power at higher rates in the crucial 1988 plebiscite that bolstered the democratic transition. Potential mechanisms include informational frictions on the intensity of repression in counties far from bases and shifts in preferences caused by increased proximity to the events. Election outcomes after democratization show no lasting change in political preferences.*

Characterize repression compliers

Table C4: Characterization of compliers

	Treated Compliers	Untreated Compliers	Full sample
	(1)	(2)	(3)
A. Pre-1973 characteristics:			
Houses per capita in 1970	0.19	0.22	0.20
Land inequality 1965 (Gini)	0.85	0.80	0.85
Agrarian reform intensity	0.10	0.24	0.20
Vote share Allende 1970	0.61	0.63	0.27
Vote share Alessandri 1970	-0.19	0.31	0.20
B. Post-1973 characteristics:			
Plebiscite:			
Registration	116.18	89.36	71.16
Vote share “No”	58.79	52.29	54.82
Repression year:			
In 1973	0.66	0.33	0.44
In 1974	0.13	0.14	0.11
≥1975	0.25	0.30	0.33
Profession:			
Laborer	0.44	0.19	0.25
Farmer	0.16	-0.08	0.09
Military	0.09	0.06	0.07

IV in R and Stata

R:

- `AER::ivreg` is the classic function
- `lfe::felm` and `fixest::feols` have built-in options for IV

Stata:

- `ivregress` is the classic function
- `ivreg2` user-written function with more estimators
- `xtivreg` and `ivreghdfe` are versions thought for panel data

As usual, check the documentation to make sure you understand what's going on (e.g. different standard error corrections)

Profile compliers in R: ivdesc

Profiling Compliers and Noncompliers for Instrumental-Variable Analysis

Moritz Marbach[1] and Dominik Hangartner[1,2]

[1] *Center for Comparative and International Studies, ETH Zurich, 8092 Zurich, Switzerland*
[2] *Department of Government, London School of Economics, London WC2A 2AE, UK. Email: d.hangartner@lse.ac.uk*

Abstract

Instrumental-variable (IV) estimation is an essential method for applied researchers across the social and behavioral sciences who analyze randomized control trials marred by noncompliance or leverage partially exogenous treatment variation in observational studies. The potential outcome framework is a popular model to motivate the assumptions underlying the identification of the local average treatment effect (LATE) and to stratify the sample into compliers, always-takers, and never-takers. However, applied research has thus far paid little attention to the characteristics of compliers and noncompliers. Yet, profiling compliers and noncompliers is necessary to understand what subpopulation the researcher is making inferences about and an important first step in evaluating the external validity (or lack thereof) of the LATE estimated for compliers. In this letter, we discuss the assumptions necessary for profiling, which are weaker than the assumptions necessary for identifying the LATE if the instrument is randomly assigned. We introduce a simple and general method to characterize compliers, always-takers, and never-takers in terms of their covariates and provide easy-to-use software in `R` and `STATA` that implements our estimator. We hope that our method and software facilitate the profiling of compliers and noncompliers as a standard practice accompanying any IV analysis.

Keywords: instrumental variables, local average treatment effects, non-compliance

Double-robust estimator and IV

Remember the Double-Robust estimator:

$$\hat{\tau}_{dr} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{D_i Y_i}{\hat{e}(X_i)} - \frac{\hat{Y}_{1i}(D_i - \hat{e}(X_i))}{\hat{e}(X_i)}\right) - \frac{1}{N}\sum_{i=1}^{N}\left(\frac{(1-D_i)Y_i}{1-\hat{e}(X_i)} + \frac{\hat{Y}_{0i}(D_i - \hat{e}(X_i))}{1-\hat{e}(X_i)}\right) =$$

$$= \frac{1}{N}\sum_{i=1}^{N}\left(\hat{Y}_{1i} - \frac{D_i(Y_i - \hat{Y}_{1i})}{\hat{e}(X_i)}\right) - \frac{1}{N}\sum_{i=1}^{N}\left(\hat{Y}_{0i} - \frac{(1-D_i)(Y_i - \hat{Y}_{0i})}{1-\hat{e}(X_i)}\right)$$

Consistent if either the specification for PS or CEF is correct. Intuition: we are correcting bias in the error of one model with an orthogonal model.

Double-robust estimator and IV

Intuition: construct DR estimators for IV.

Double-robust estimator and IV

Intuition: construct DR estimators for IV.

$$\hat{\tau}_{wald} = \frac{\hat{E}[Y_i|Z=1] - \hat{E}[Y_i|Z=0]}{\hat{E}[D_i|Z=1] - \hat{E}[D_i|Z=0]}$$

Both the numerator (ITT/RF) and the denominator (1S) are Differences-in-means which identify causal effects, with a common treatment Z.

One can build DR estimators for both and, if the estimated PS function for the instrument Z is correctly specified, have consistency in both. Then, the ratio is consistent too (e.g. Slutsky theorem)