Intermediate Econometrics

26th November 2025 - Vincenzo Gioia

An introduction to causal inference

Instrumental Variable Regression

- Use of instrumental variables in the context of the potential outcome model
- Focus on the simplest case:

1. treatment d is binary
2. instrument z is binary (setting of the Wald estimator)

- $y_i(z_i, d_i)$: value of the outcome for the i-th individual for a given combination of z and d

Introduction to causal inference

IV Regression Assumptions : Exclusion Restriction

1. z can be used as an instrumental variable if z is related to y only because of its influence on d, that is for a given value of d, the outcome y is the same whatever the value of z (**exclusion restriction**):

$$y_i(z_i = 0, d_i) = y_i(z_i = 1, d_i)$$

- Then, potential outcome can be defined as a function of the treatment variable only: $y_i(z_i, d_i) = y_i(d_i)$
- Further, d_i can be expressed as a function of z_i: $d_i(z_i)$.

Introduction to causal inference

IV Regression Assumptions: Causal effect and monotonicity

2. On average, z has a **causal effect** on d:

$$\mathrm{E}\left(d(1) - d(0)\right) \neq 0$$

.

3. **Monotonicity**:

$$d_i(1) \geq d_i(0) \quad \forall i$$

Introduction to causal inference

IV regression: Results under the assumptions

- With the assumptions above, the causal effect of z on y is:

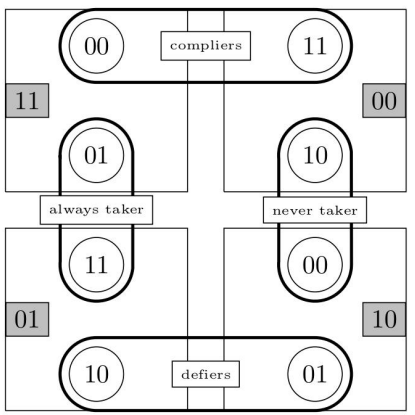
$$\begin{aligned} y_i(z_i = 1, d_i(1)) - y_i(z_i = 0, d_i(0)) &= y_i(d_i(1)) - y_i(d_i(0)) \\ &= [y_i(1)d_i(1) + y_i(0)(1 - d_i(1))] \\ &- [y_i(1)d_i(0) - y_i(0)(1 - d_i(0))] \\ &= (y_i(1) - y_i(0))(d_i(1) - d_i(0)) \end{aligned}$$

- The causal effect of z on y is the product of the causal effects of d on y and of z on d

Introduction to causal inference

IV regression: Results under the assumptions

- Consider now the relation between z and d at the individual level. There are four observable categories of individuals (indicated in a gray square) that can be considered
- Counterfactuals are represented by two circles inside the square and we can distinguish:
- **compliers**: $d_i(1) = 1$ and $d_i(0) = 0$
- **always takers**: $d_i(1) = d_i(0) = 1$
- **never takers**: $d_i(1) = d_i(0) = 0$
- **deniers**: $d_i(1) = 0$ and $d_i(0) = 1$

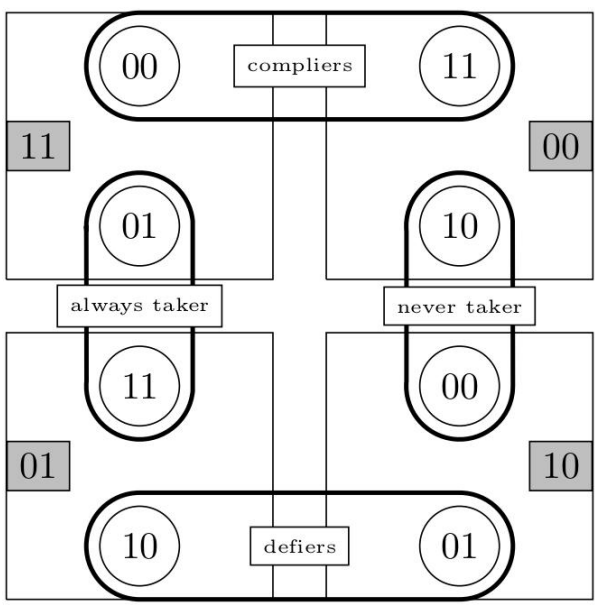


Introduction to causal inference

IV regression: Results under the assumptions

- For example, the square called 00 contains the individuals for which $z = 0$ and $d = 0$:

1. For those individuals the unobserved counterfactual is $z = 1$ and d either equal to 0 or 1
2. If the counterfactual is $z = 1, d = 0$, the individual is a **never-taker**, $d = 0$ whatever the value of z
3. If the counterfactual is $z = 1, d = 1$, the individual is a **complier**



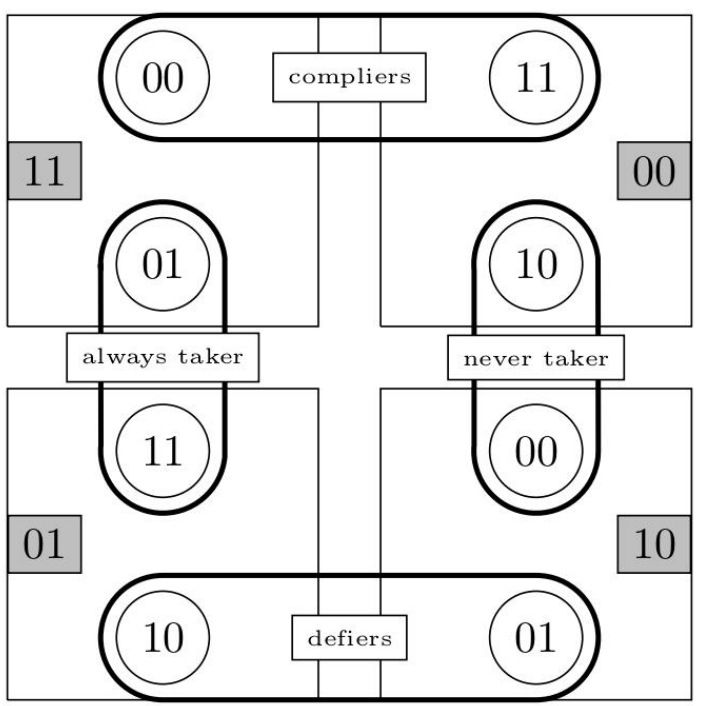
Introduction to causal inference

IV regression: Results under the assumptions

- *The monotonicity assumption hypothesis rules out the existence of deniers*
- **monotonicity**:

$$d_i(1) \geq d_i(0) \quad \forall i$$

- **deniers**: $d_i(1) = 0$ and $d_i(0) = 1$

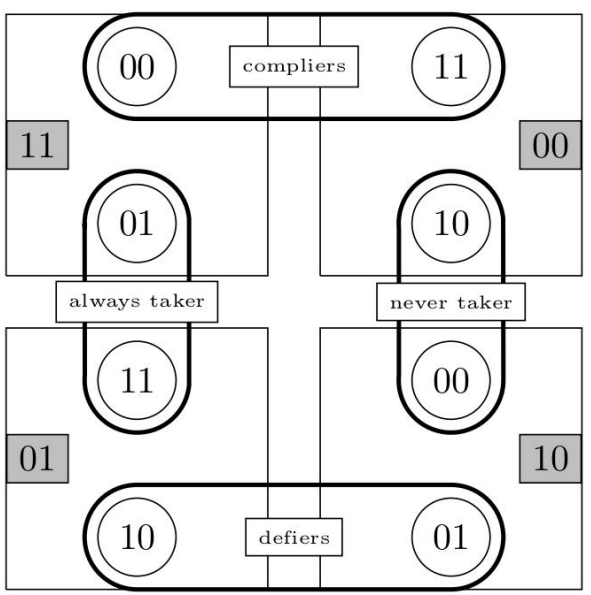


Introduction to causal inference

IV regression: Results under the assumptions

- *With the always takers and the never takers, the causal effect of z on y in is zero because the causal effect of z on d is zero*
- **always takers:** $d_i(1) = d_i(0) = 1$
- **never takers:** $d_i(1) = d_i(0) = 0$

$$y_i(z_i = 1, d_i(1)) - y_i(z_i = 0, d_i(0)) = (y_i(1) - y_i(0))(d_i(1) - d_i(0)) = 0$$

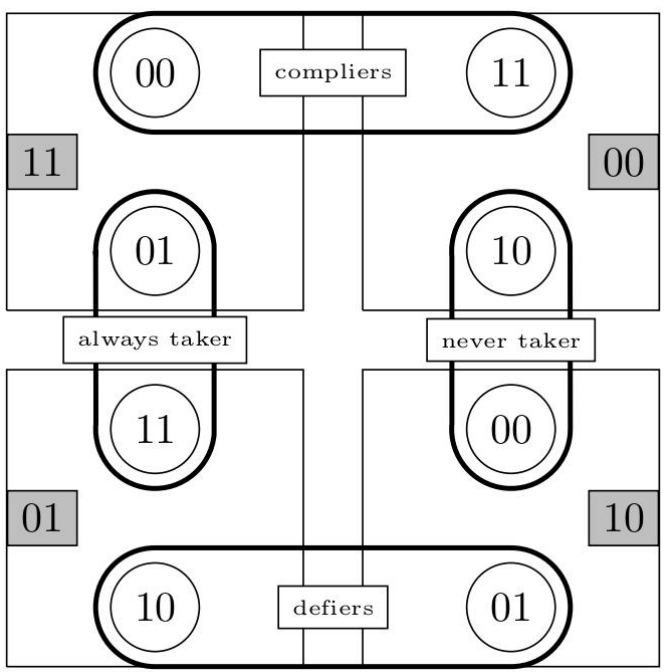


Introduction to causal inference

IV regression: Results under the assumptions

- **Therefore, the causal effect of z on d reduces to the treatment effect for the compliers**
- **compliers**: $d_i(1) = 1$ and $d_i(0) = 0$

$$y_i(z_i = 1, d_i(1)) - y_i(z_i = 0, d_i(0)) = (y_i(1) - y_i(0))(d_i(1) - d_i(0))$$



Introduction to causal inference

IV regression: Results under the assumptions

- The IV estimator, allows to estimate, in the general case, the following population estimand: $\frac{\mathbf{COV}(y,z)}{\mathbf{COV}(d,z)}$, which, for the case where both d and z are binary, reduces to:

$$\mathbf{LATE} := \frac{\mathrm{E}\left(y(1, d(1)) - y(0, d(0))\right)}{\mathrm{E}\left(d(1) - d(0)\right)}$$

1. Numerator: the average causal effect of z on y for the compliers
2. Denominator: the share of compliers in the population

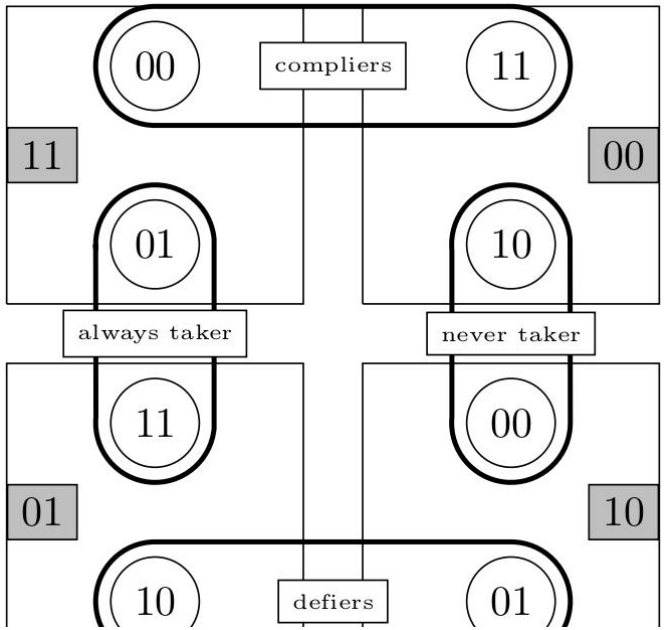
- LATE is the acronym of **local average treatment effect**, stressing the fact that what is estimated using the instrumental variable estimator **is an average treatment effect for only a subset of the population called the compliers**, that is those for which a change in the value of z results in a change in the value of d

Introduction to causal inference

IV regression: Results under the assumptions

$$\textbf{LATE} := \frac{\mathrm{E}\left(y(1, d(1)) - y(0, d(0))\right)}{\mathrm{E}\left(d(1) - d(0)\right)}$$

- The numerator is also called the **reduced form** equation, as it measures the effect of the instrument on the outcome
- The denominator is called the **intention to treat** equation, it measures the effect of the instrument on the treatment dummy



Introduction to causal inference

Econometrix example

- Dataset `paces`, extracted from Angrist (1992)
- Goal: investigate the effect of a large school voucher program in Columbia called PACES
- This voucher covers more than half of the cost of private secondary school and may induce parents to enroll their children in private schools, which are known to provide much better service than public schools

```
1 library(micsr)
2 paces <- as.data.frame(micsr.data::paces)
```

Introduction to causal inference

Econometrix example

- Here, z is a dummy for children who receive the voucher (**voucher**) and d is a dummy for enrollment in private school (tretment variable, **privsch**)
- We will consider as outcome only the number of years of finished education **educyrs**
- The OLS estimate of the treatment is just the difference between the mean of the outcome for the two subsamples defined by the treatment variable
- Then, the number of years of education is higher by about 0.29 of a year for pupils enrolled in private schools, and the effect is highly significant

```r
ols <- lm(educyrs ~  privsch, data = paces)
summary(ols)$coefficients
```

```
             Estimate Std. Error    t value     Pr(>|t|)
(Intercept) 7.2288786 0.04270619 169.270027 0.000000e+00
privsch     0.2916397 0.05573156   5.232936 1.892721e-07
```

Introduction to causal inference

Econometrix example

- Let's see the effect of the instrument on the treatment (intention to treat) and on the outcome (reduced form)
- The voucher has a large effect on enrolling in private school, the intention to treat effect being 0.64. The effect of the instrument on the outcome is 0.108

```
intTreat <- lm(privsch ~ voucher, data = paces)
summary(intTreat)$coefficients
```

```
             Estimate Std. Error  t value      Pr(>|t|)
(Intercept) 0.2549277 0.01354494 18.82089  1.946945e-71
voucher     0.6421311 0.01882989 34.10169 2.386245e-191
```

```
reduced <- lm(educyrs ~ voucher, data = paces)
summary(reduced)$coefficients
```

```
            Estimate Std. Error    t value   Pr(>|t|)
(Intercept) 7.344284 0.03979311 184.561685 0.00000000
voucher     0.107922 0.05531955   1.950884 0.05124794
```

Introduction to causal inference

Econometric example

- The IV estimator is the ratio of these two effects, which is 0.168 (to be compared to 0.29 of the OLS regression)
- Therefore, in this example, we get an IV estimator of the treatment effect much smaller than the OLS estimator, which may be the symptom that unobserved determinants of the outcome are positively correlated with the enrollment in private school

```
coef(reduced)[2]/coef(intTreat)[2]
```

```
  voucher
0.1680686
```

```
iv1 <- lm(educyrs ~  predict(intTreat), data = paces)
iv1$coefficients[2]
```

```
predict(intTreat)
        0.1680686
```

```
library(ivreg)
iv2 <- ivreg(educyrs ~  privsch | voucher, data = paces)
iv2$coefficients[2]
```

```
  privsch
0.1680686
```

Introduction to causal inference

Econometric example

- Covariates can easily be added to the analysis. The authors of the study use the covariates

1. **pilot**: (=1) if the individual was surveyed during the *pilot* survey;
2. **housvisit**: (=1) if the survey was conducted in person and not by phone
3. **smpl**: a factor indicating the three subsamples (Bogota in 1995, Bogota in 1997 and Djamunid in 1993)
4. **phone**: a dummy for owning a phone
5. **age**: age of the pupil
6. **sex**: sex of the pupil
7. **strata**: a strata of residence

Introduction to causal inference

Econometric example

- We then compute the OLS and IV regression
- The two estimates are now almost identical: introducing the covariates reduces by almost one half the value of the OLS estimator and has a small effect on the IV estimator without covariates
- After controlling for covariates, the omitted variable bias is negligible

```r
ols <- lm(educyrs ~ privsch + pilot + housvisit + smpl +
                phone + age + sex + strata + month, data = paces)
iv <- ivreg(educyrs ~ privsch + pilot + housvisit + smpl +
                      phone + age + sex + strata + month |
                      voucher + pilot + housvisit + smpl +
                      phone + age + sex + strata + month, data = paces)
rbind(summary(ols)$coefficients[2,], summary(iv)$coefficients[2,])
```

```
     Estimate Std. Error  t value     Pr(>|t|)
[1,] 0.1408088 0.04239104 3.321663 0.0009156163
[2,] 0.1342252 0.06510330 2.061726 0.0393999944
```

Introduction to causal inference

A note on Regression discontinuity

- Eligibility to a program is sometimes based on the value of an observable variable (called the **forcing variable**)
- More precisely, on the fact that the value of this variable, for an individual is below or over a given threshold
- Individuals just below and just over the threshold therefore constitute two groups of individuals who are very similar, except that the first group receives the treatment and the second group doesn't. This is called a **regression discontinuity** (**RD**) design
- Two variants of regression discontinuity designs can be considered:

1. **sharp discontinuity**, there is a one-to-one correspondence between eligibility and treatment
2. **fuzzy discontinuity**, the probability of being treated is very different just below and just over the threshold

Introduction to causal inference

Difference-in-Differences

- Sometimes, the outcome is observed for two periods:

1. In the first period, the treatment hasn't been implemented
2. For the second period, some individuals have been treated (the treatment group), as some individuals haven't (the control group).

- The effect of the treatment can then be estimated by:

$$\frac{\sum_{i=1}^{n_T}(y_{i2} - y_{i1})}{n_T} - \frac{\sum_{i=1}^{n_C}(y_{i2} - y_{i1})}{n_C}$$

- Each term is the mean difference of the outcome between the two periods for the two groups, and the estimator is the difference of these differences

Introduction to causal inference

Difference-in-Differences: Theoretical Example

- Let's assume that the treatment is a job-training program implemented in 2021 and that the outcome is wage observed in 2022
- If the first and second term of

$$\frac{\sum_{i=1}^{n_T}(y_{i2} - y_{i1})}{n_T} - \frac{\sum_{i=1}^{n_C}(y_{i2} - y_{i1})}{n_C}$$

are \$1000 and \$600, this means

1. that the average annual wage in the treatment group increased by \$1000 in 2022 compared to 2021. This would be a relevant estimator of the effect of the program if nothing had changed on the labor market in 2022 compared to 2021
2. that the average annual wage in the control group increased by \$600 (if the economic situation improved, then the average wages will increase even for those who haven't been treated)

- **Therefore, the effect of the treatment is the difference between the two terms, which is \$400**

Introduction to causal inference

Difference-in-Differences: Practical Example

- Let's suppose to estimate the causal effect of police on crime
- This task is difficult using non-experimental data, as there may be a **reverse causality relationship** between police and crime:

1. **more police reduces crime (negative causal relationship)**
2. **but an increase in crime may lead authorities to increase police (positive reverse causal relationship)**

- In July 18, 1994, a terrorist attack destroyed the main Jewish-owned center in Buenos Aires, Argentina, killing 85 people, and the federal government decided to provide 24 hour police protection to Jewish-owned institutions

Introduction to causal inference

Econometric example

- Study of Di Tella and Schargrodsky (2004)
- Data on three “barrios” in Buenos Aires at the block level on a monthly basis
- The outcome of interest is the number of car thefts
- Denoting with x_1 a dummy equal to 1 if the block contains or is close to a Jewish institution and x_2 a dummy equal to 1 after the attack
- The treatment effect is the estimate of β_4 on the following equation:

$$y_{it} = \beta_1 + \beta_2 x_{i1} + \beta_3 x_{it2} + \beta_4 x_{i1} x_{it2} + \epsilon_{it}$$

with $t = 1, 2$, (the periods before and after the attack).

Introduction to causal inference

Econometric example

- The `car_thefts` data set contains repeated observations of car thefts (**thefts**) for 876 blocks.
- Each block is observed 10 times on a monthly basis, and the month of the attack (July) is split into two half-month observations
- We first compute the total number of thefts (`thefts`) before and after the attack for all the blocks
- As the two periods are of unequal length (3.5 and 4.5 months), we divide the number of thefts for the two periods by the corresponding number of days and we multiply by 30.5 to get a monthly value: **The number of monthly car thefts per block is about 0.09**

```
car_thefts <- as.data.frame(micsr.data::car_thefts)

sum_thefts <- aggregate(thefts ~ block + period, data = car_thefts, sum)
sum_days <- aggregate(days ~ block + period, data = car_thefts, sum)
two_obs <- merge(sum_thefts, sum_days, by = c("block", "period"))
two_obs$thefts <- two_obs$thefts / two_obs$days * 30.5
mean(two_obs$thefts)
```

```
[1] 0.09327905
```

Introduction to causal inference

Econometric example

- **distance** is a factor indicating the distance from the block to the nearest Jewish-owned institution: its levels are “same” (same block), “one”, “two” and “>2” (one block, two blocks or more than two blocks).
- We add this variable to new **two_obs** data.frame by selecting distance and block in the original data frame, selecting only the distinct rows (one per block) and joining it to two_obs

```
block_distance <- unique(car_thefts[, c("block", "distance")])
two_obs <- merge(two_obs, block_distance, by = "block", all.x = TRUE)
```

Introduction to causal inference

Econometric example

- We then compute the regression by coercing the `distance` to a dummy for the same block
- The effect of police on thefts is -0.07, which is very high as the average number of monthly theft per block is 0.09 and it is significant

```
two_obs$distance <- ifelse(two_obs$distance == "same", 1, 0)
mod <- lm(thefts ~ period * distance, data = two_obs)
summary(mod)$coefficients
```

```
                          Estimate  Std. Error    t value     Pr(>|t|)
(Intercept)            0.089150917 0.004558794 19.5558136 3.575037e-77
periodafter            0.010551986 0.006447108  1.6367008 1.018730e-01
distance               0.008982932 0.022182021  0.4049645 6.855531e-01
periodafter:distance -0.072318585 0.031370115 -2.3053337 2.126450e-02
```

Introduction to causal inference

Econometric example

- The same difference-in-differences estimator can be obtained by computing a t-test of equality of the two means
- We first reshape the data set in order to have one line per block, and we compute the after-before difference:

```r
before <- two_obs[two_obs$period == "before",
                  c("block", "distance", "thefts")]
after  <- two_obs[two_obs$period == "after",
                  c("block", "distance", "thefts")]

names(before)[names(before) == "thefts"] <- "before"
names(after)[names(after)   == "thefts"] <- "after"

diffs <- merge(before, after, by = c("block", "distance"))

diffs$dt <- diffs$after - diffs$before
```

Introduction to causal inference

Econometric example

- The same difference-in-differences estimator can be obtained by computing a t-test of equality of the two means

```
mean(diffs$dt[diffs$distance == 1]) - mean(diffs$dt[diffs$distance == 0])
```

```
[1] -0.07231858
```

```
t.test(dt ~ factor(distance), diffs, var.equal = TRUE)
```

```
    Two Sample t-test

data:  dt by factor(distance)
t = 2.7388, df = 874, p-value = 0.006291
alternative hypothesis: true difference in means between group 0 and group 1 is not
equal to 0
95 percent confidence interval:
 0.02049419 0.12414298
sample estimates:
mean in group 0 mean in group 1
     0.01055199     -0.06176660
```

Introduction to causal inference

Econometric example

- The difference-in-differences can be extended to the case where the observation units before and after the implementation of the treatment are not the same
- Let's consider the case in Hong (2013), which studied the impact of the introduction of Napster on music expenditure
- The study is based on the Consumer Expenditure Survey, which is performed on a quarterly basis, and the data set is called `napster`
- Napster was introduced in June 1999 and became the dominant file-sharing service
- Households with (without) internet access constitute the treatment (control) group

Introduction to causal inference

Econometric example

- From the `date` series, we construct a `period` variable using June 1999 as the cutoff:

```
napster <- as.data.frame(micsr.data::napster)

napster <- napster[, c("date", "expmusic", "internet", "weight")]
cutoff <- as.Date("1999-06-01")

napster$period <- ifelse(as.Date(napster$date) < cutoff,
                         "before",
                         "after")

napster$period <- factor(napster$period, levels = c("before", "after"))
```

Introduction to causal inference

Econometric example

- We then proceed to the estimation, with expmusic, the expenditures on recorded music as a response
- `internetyes` has a strong positive effect on expenditure. The interaction term between `period` and `internet` indicates that, the deployment of Napster led to a significant reduction of the expense for the “treated” individuals (those who have internet access) of $4.6.

```
1  fit <- lm(expmusic ~ period * internet, napster, weight = weight)
2  summary(fit)$coefficients
```

```
                          Estimate Std. Error   t value      Pr(>|t|)
(Intercept)              10.560788  0.1749405 60.367880  0.000000e+00
periodafter              -1.749118  0.2532558 -6.906529  4.993947e-12
internetyes              14.781168  0.4318063 34.231013 1.992747e-255
periodafter:internetyes -4.588916  0.5517472 -8.317063  9.120641e-17
```

Introduction to causal inference

Matching and Propensity Score Matching

- Matching is used to make treated and control units comparable in observational data.
- When covariates are many or continuous, exact matching becomes impossible
- This is why we use the **propensity score matching**, the estimated probability of receiving the treatment
- However, estimating the propensity score requires understanding logistic (or probit) regression first, that we will introduce on friday

Sinthetic control Methods

- A transparent, data-driven counterfactual for single-unit interventions

1. Used when treatment affects **one aggregate unit** (e.g., a region or country).
2. Builds a **synthetic version** of the treated unit using weighted donor units.

- Key idea: Choose weights so the synthetic unit **matches pre-treatment trends** and compute the post-treatment gaps (**causal effect**)
- Example: Basque Country terrorism: synthetic control = **85% Catalonia + 15% Madrid**

Introduction to causal inference

References

- Angrist, J.D., Bettinger, E., Bloom, E., King, E., and Kremer, M. (2002) “Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment.” American Economic Review 92 (5): 1535–58.
- Di Tella, R., and Schargrodsky, E. (2004). “Do Police Reduce Crime? Estimates Using the Allocation of Police Forces After a Terrorist Attack.” The American Economic Review 94 (1): 115–33
- Hong, Seung-Hyun (2013) “Measuring the Effect of Napster on Recorded Music Sales: Difference-in-differences Estimates Under Compositional Changes”, Journal of Applied Econometrics, 28, 297-324