Final Fixed Effects

Tytus Wilam

12/13/2019

(1) Hypothetical real-life scenario.

The motivation for the study is to find out if online ads work better on some people than on others. Fixed effects allow us to capture the heterogeneity of our subjects without measuring it.

This is a study on the causal effects of online search ads on consumer behavior. Each year for ten years we have collected data from 100 working adults who are using an online search engine. The treatment (z) is visual exposure to a search engine ad. The outcome (y) is the purchase of advertised product. We know about the following covariates for each person for each year: hourly wage, is the person married? (binary), did the person re-locate? (binary). The hypothetical benefit of the study would be to evaluate if the ads work and t oimprove ad targeting in the future. In World B we also measure a mediating variable that is affected by the treatment and affects the outcome: wether the person clicked on the advertisement.

(2) DGP.

I generate panel data for 100 study participants. Hourly income for first year of the study is distributed normally and rises by 2 every year:

$$Income \sim N(30, 10)$$

$$income_{year} = income_0 + year * 2$$

Where year indexes years from the beginning of the study. Each participant begins unmarried and is equally likely to get married in any year of the study.

Each participant has a 0.2 chance to relocate each year.

Probability of a positive outcome, i.e. buying the product and y=1, is a function of the covariates for each person and for each year, and of the treatment effect of 0.2. The functions for World A and World B respectively are the following:

$$p_{year} = 0.1 + income_{year} * 0.005 + married_{year} * 0.1 + relocated_{year} * 0.2 + z * 0.2$$

$$p_{year} = 0.1 + clicked_{year} * 0.25 + married_{year} * 0.1 + relocated_{year} * 0.2 + z * 0.2$$

In World B each participant in each year they were selected for treatment (z=1) has a 0.7 probability of clicking on the ad they were shown:

$$P(clicked_{iyear} | z_{iyear} = 1) = 0.8$$

(3) DGP code.

```r
### (3)(i) DGP code for World A
set.seed(1234)
participant <- 1:100
df <- tbl_df(data.frame(id=NA, year=NA, income = NA, married = NA,
                        relocated=NA, z=NA, y0=NA, y1=NA, y=NA))

for (i in participant) {
  id <- rep(i, 10)
  year <- 1:10
```

```
  income <- rnorm(1, 30, 10) + 1:10 * 2
  temp <- sample(1:10, 1)
  married <- c(rep(0, temp), rep(1, 10-temp))
  relocated <- rbinom(10, 1, 0.2)
  z <- rbinom(10, 1, 0.5)
  prob <- 0.1 + income*0.005 + married*0.1 + relocated*0.2
  y0 <- rbinom(10, 1, prob)
  y1 <- rbinom(10,1, prob+0.2)
  y <- ifelse(z==1, y1, y0)
  table <- tbl_df(data.frame(id, year, income, married, relocated, z, y0, y1, y))
  df <- bind_rows(df, table)
}

df <- na.omit(df)
worldA <- df
worldA.obs <- select(worldA, id:z, y)
```

```
### (3)(ii) DGP code for World B
set.seed(1234)
participant <- 1:100
df <- tbl_df(data.frame(id=NA, year=NA, relocated = NA, clicked = NA, married = NA, z=NA, y0=NA, y1=NA,
for (i in participant) {
  id <- rep(i, 10)
  year <- 1:10
  temp <- sample(1:10, 1)
  married <- c(rep(0, temp), rep(1, 10-temp))
  relocated <- rbinom(10, 1, 0.2)
  z <- rbinom(10, 1, 0.5)
  clicked <- ifelse(z==1, rbinom(1,1,0.8), 0)
  prob <- 0.1 + clicked*0.25 + married*0.1 + relocated*0.2
  y0 <- rbinom(10, 1, prob)
  y1 <- rbinom(10,1, prob)
  y <- ifelse(z==1, y1, y0)
  table <- tbl_df(data.frame(id, year, clicked, married, relocated, z, y0, y1, y))
  df <- bind_rows(df, table)
}

df <- na.omit(df)
worldB <- df
worldB.obs <- select(worldB, id:z, y)
```

(4)(a) Description of method and the estimand

I use the fixed effects approach. The estimand is the weighted average of within-individual treatment effects. I am able to do so because I have within-individual empirical counterfactual for all observations.

I perform a regression on the outcome that includes treatment, the covariates, and an indicatopr variable (id) for each individual. The indicator variable captures individual specific confounders. individual-specific confounders. This should yield results that are not biased on average.

```
### (4)(b)(i) Code used to estimate the results in World A. (I show the results in 6.)
estA <- summary(lm(y ~ z + factor(id) + income + married + relocated, worldA.obs))$coef
# The treatment z causally increases the purchases by 16.64 %
```

```r
### (4)(b)(ii) Code used to estimate the results in World (B). (I show the results in 6.)
estB <- summary(lm(y ~ z + factor(id) + relocated + married + clicked, worldB.obs))$coef
```

(5) Assumptions

I assume that the characteristics that cannot be observed are fixed and time-invariant for each individual. The entity-specific constant alpha is constant. I take advantage of that and find a unique coefficient that captures the unobserved for each individual. It also makes it possible for me to estimate unbiased slope coefficients for the covariates. I am only able to do that because I have panel data which gives me multiple observations of each person.

Perhaps more importantly, I assume ignorability also known as "all confounders measured." Ignorability means that there are no unmeasured variables that may be associated both with the treatment and the outcome and that only confounders are measured.

$$Y(1), Y(0) \perp Z | X, \alpha$$

Where α indicates individual fixed effects.

I also make a parametric assumption that a linear regression model is appropriate.

(6) Attractive display of results.

```
## [1] "World A"
```

```
##              Estimate  Std. Error    t value     Pr(>|t|)
## z          0.16644902 0.032314492  5.1509093 3.188427e-07
## income     0.01028530 0.003752282  2.7410794 6.245681e-03
## married   -0.02155799 0.053088234 -0.4060785 6.847819e-01
## relocated  0.17841073 0.040756623  4.3774660 1.342160e-05
```

```
## [1] ""
```

```
## [1] "World B"
```

```
##             Estimate Std. Error   t value     Pr(>|t|)
## z         0.03323428 0.05553414 0.5984478 5.496924e-01
## relocated 0.18813046 0.03544548 5.3076016 1.401027e-07
## married   0.10648078 0.03270157 3.2561366 1.171785e-03
## clicked   0.19178771 0.06428304 2.9834884 2.927130e-03
```

In World A the weighted within-individual average causal effect is estimated to be 17%. This means that an individual treated with an advertisement is 17% more likely to have an outcome of buying the product in a given year compared to the counterfactual scenario in which they weren't exposed to the advertisement.

$$Y(Z = 1) - Y(Z = 0) = 0.17$$

(7) Bias and assumption violation.

In world A I did not violate any assumptions and my estimated is mostly unbiased; fixed effects work as they are supposed to.

In World B I falsely assume:

$$Y(1), Y(0) \perp Z | married, relocated, clicked, \alpha$$

I include the "clicked" covariate which is a mediating variable between the treatment and the outcome. The real world scenario is that google advertisements are supposed to work mostly by getting people to click on them. Therefore, "clicked" is both influenced by the treatment and influences the outcome. In fact, according to the DGP, seeing the advertisement (treatment z) affects the outcome only by facilitating clicking.

The estimate for world B is biased. The fixed effect estimand in world B is 0.03. What does it mean? The weighted average of within-individual causal effects negligible. The causal interpretation of this number is that if an individual is exposed to an advertisement in that year he is 3% more likely to purchase the advertised product compared to the counterfactual scenario in which they are not exposed to the treatmment. This is far from the true causal effect of 20%, which is entirely captured by the clicked covariate. This effect disappears entirely when I do not control for clicked.

The results for the world B are biased because I violated ignorability by controlling a variable that was a mediating variable instead of for a confounder. This need not be a problem in a different research design. Measuring the causal effect of "clicked" is a useful research project on its own. However, with this design the researcher will mistakenly conclude that the treatment has little effect on the outcome instead of seeing that the effect is shared between the treatment and the mediating variable.

(8) Lessons learned I learned that: (a) One must not control for a covaraite that is affected by the treatment and affects the outcome. If I control for it, I will get a mistaken estimate of the causal effect of the treatme t because the treatment works through that mediating variable. (b) Fixed effects can be used to eliminate individual-specific confounders without measuring them.

I have also learned that working with panel data is difficult. Initially, I tried to simulate the study we talked about in class where panel data was used to estimate the effect of marriage on men's earnings. That study controlled for post-treatment variables and I tried to figure out a way to conduct it that would not violate any assumptions. I was not able to do that.