UCLAAnderson
SCHOOL of MANAGEMENT

PMDA

Lecture 4: Control Variables

Upcoming

- Problem Set # 2: Feb 11, 11.59 pm

Recap from last class: A/B testing

- Understand the power of randomization
    - Simple method to make causal statements → reliable, scalable, transparent, easy to communicate
    - Only important issue: randomly assign treatment variable
- Learn how to run more precise A/B tests
- Three ways to affect precision
    - Maximize Var(X) by splitting sample equally into treatment/control
    - Reduce variance of regression residual through control variables
    - Choose sample size (after optimizing along other two margins)

How to increase precision in A/B tests in practice

1. Determine the SE you need in order to reject the null hypothesis. For a positive ROI in the example where an ad costs $0.04, the hypothesis is
$$H_0 : \beta_1 = 0.04$$
$$H_1 : \beta_1 > 0.04$$
2. Pick a confidence level, e.g. 97.5% (which gives you the easy-to-use critical value 1.96 for a one-sided test)
3. You reject if $(\hat{\beta}_1 - 0.04)/SE > 1.96$ which means $SE < (\hat{\beta}_1 - 0.04)/1.96$. So, given a preliminary estimate of $\hat{\beta}_1$ choose a SE that satisfies this condition
4. Choose p = prob of treatment to maximize the variance of the treatment variable X. Since for a binary variable Var(X) = p*(1-p) → largest if we choose p = 0.5
5. Get a preliminary estimate of s^2 by running a multivariate regression with as many controls as you can (we will see later why this makes sense)
6. Given the chosen SE, Var(X) and s^2, choose the sample size as
$$N = \frac{s^2}{Var(X)} \frac{1}{SE^2}$$

Overview for today

- Next three lectures:
  - If you can't run experiments, can you still make causal claims in observational data?
  - This week we will see how sometimes you can use controls to isolate random variation in data that is not fully randomized

- Key idea today:
  - X-variable might vary for many reasons, some of them random
  - Multivariate regression can be used to isolate the random part of the variation in X and remove the non-random part

- Goals:
  - Understand partial randomization
  - Econometric concepts: sequential estimation of multivariate regression + dummy variables with multiple categories

Agenda today

- Econometric theory: sequential estimation of multivariate regression and "residual variation"
- Example: targeted advertising
- Workshop

Estimation of multivariate regression

- Recall univariate regression (where “i” is the unit of observation in the data)

$$Y_i = \beta_0 + \beta_1 X_i + e_i$$

- We estimate the coefficients by minimizing the sum of squared errors

$$\min_{\beta_0,\beta_1} \sum_i (e_i)^2 = \min_{\beta_0,\beta_1} \sum_i (Y_i - (\beta_0 + \beta_1 X_i))^2$$

- This is also how we estimate the β ‘s in multivariate regressions

$$Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki} + e_i$$

$$\min_{\beta_0,\beta_1,\ldots,\beta_k} \sum_i (e_i)^2 = \min_{\beta_0,\beta_1,\ldots,\beta_k} \sum_i (Y_i - (\beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki}))^2$$

Sequential estimation of multivariate regression

- Let's drop the "i" subscript

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + e$$

- Frisch-Waugh Theorem: In principle, we could estimate each β sequentially. E.g.,
  1. First regress X_1 on all other X-variables

$$X_1 = a_0 + a_1 X_2 + a_2 X_3 + \cdots + \widetilde{X}_1$$

  2. Then regress Y on the residual $\widetilde{X}_1$ from the regression in step 1

$$Y = \beta_0 + \beta_1 \widetilde{X}_1 + v$$

The coefficient on the residual in the regression in step 2 equals β_1

- In practice, we estimate all the β's jointly not sequentially, but the theorem helps understand what a multivariate regression does in the background

Usefulness of Frisch-Waugh theorem/1

- We used this theorem last week to understand why control variables improve the precision of the treatment effect estimator under full randomization

- Recall that we looked at the variance of the treatment effect estimator in the multivariate regression, which was

$$Var(\hat{\gamma}_{1,multi\_var}) = \frac{1}{N} \frac{\tilde{\sigma}^2}{Var(\widetilde{treatment})}$$

- In this week's notation from the previous slide, $\hat{\gamma}_{1,multi\_var}$ is the estimator of β_1, $\widetilde{treatment}$ is $\widetilde{X}_1$ and $\tilde{\sigma}^2$is the variance of e

→ Frisch-Waugh theorem gives an expression for the precision of the coefficient estimators in multivariate regressions that is based on sequential estimation

Usefulness of Frisch-Waugh theorem/2

- The Frisch-Waugh theorem gives intuition for why controls remove variation in X_1 that in a univariate regression would cause omitted variable bias
- The theorem again:

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + e$$

1. First regress X_1 on all other X-variables
2. Then regress Y on the residual from the regression in step 1

- Multivariate regression thus gives you the effect of X_1 on Y after removing the variation in X_1 that can be explained by the other X -variables
- **Q:** In the red meat example, what part of red meat consumption is eliminated by the multivariate regression, in words?
- **A:** The variation in red meat consumption that is explained by smoking and other lifestyle choices. This is the part that would cause omitted variable bias in the univariate regression

Usefulness of Frisch-Waugh theorem/3

- When X_1 is a treatment variable in experimental data (e.g., seeing an ad or not), we now see how sometimes you can use control variables to remove endogeneity caused by treatment that was not fully randomly assigned

- You can do this when you know that the treatment was not random because it was assigned based on some observable characteristics

- We are thus considering a new scenario where you have experimental data but you know that the treatment was not fully randomized, only partially randomized
  - Online targeted ads are a prime example of this that we now discuss in detail

Agenda today

- Econometric theory: sequential estimation of multivariate regression and “residual variation”
- Example: targeted ads
- Workshop

Load data & some background

- Load the dataset “targeted_ads.csv” into Jupyter Notebook
- Data contains information from an ad experiment
  - Ads (“treatment”) were not fully randomized
  - More ads were shown to younger users because they generated more revenue for the company in the past → targeting based on age/past revenue
- The company used a specific process to serve ads:
  - Users were divided into 3 age categories (bins): < 25, 25-40 and > 40
  - When a user arrived to the webpage they were shown an ad with a specific probability
  - Probability was higher for younger age bins
  - This means that treatment assignment was partially random (random within age bins but not random across age bins)

Check for random assignment

- First thing we always do with experimental data is to check if treatment (seeing the ad) was randomly assigned

Q: What regression would you run to check randomization, based on what you know about how the company assigned ads?

A: See if treatment is correlated with age. In this case, all we know is which of 3 age bins each user belongs to, so we estimate the regression

$$treatment = \beta_0 + \beta_1 age_{25to40} + \beta_2 age_{above40} + e$$

Detour: dummy variables with multiple categories

$$treatment = \beta_0 + \beta_1 age_{25to40} + \beta_2 age_{above40} + e$$

Q: Why didn't we include the third bin, age_below25, in the regression?

What happens when you run a regression with dummies for all three age bins?

A:

- A regression with all dummies is not feasible (it is an example of perfect multicollinearity)
- If you add all dummies you are trying to solve an illogical question
- Regardless of the number of categories, with dummy variables you must always leave out one category (you choose which one)

Detour: interpreting coefficients for multiple categories

$$treatment = \beta_0 + \beta_1 age_{25to40} + \beta_2 age_{above40} + e$$

Q: What is the interpretation of the coefficients in the regression when dummies have multiple categories?

A:

- β_0 is probability of treatment for omitted category age_below25
- β_1 is additional probability for age_25to40 relative to omitted category
- β_2 is additional probability for age_above40 relative to omitted category

Back to example: check non-random assignment

Q: Regress treatment on age_25to40 and age_above40. Interpret the regression coefficients. What is the probability of treatment for each age bin?

	coef	std err	t	P>|t|
Intercept	0.7508	0.022	34.788	0.000
age_25to40	-0.5542	0.030	-18.671	0.000
age_above40	-0.6466	0.030	-21.558	0.000

A:

- 75% probability for users <25 years of seeing the ad
- 55 percentage points lower probability of seeing the ad for 25 to 40 year olds relative to youngest users (i.e. 20% probability of seeing ad)
- 65 percentage points lower probability of seeing the ad for users above 40 relative to youngest users (i.e. 10% probability of seeing ad)

Back to example: check non-random assignment

Q: Is the treatment assignment random across ages? How to use the regression output to test this hypothesis?

	coef	std err	t	P>|t|
Intercept	0.7508	0.022	34.788	0.000
age_25to40	-0.5542	0.030	-18.671	0.000
age_above40	-0.6466	0.030	-21.558	0.000

A:

- Random assignment would imply that these probabilities are all equal to each other
- The p-values tell you that both age_25to40 and age_above40 are significantly different from the omitted bin age_below25
- So we reject the null that all three probabilities are equal (it's enough to find a pair that is not equal to reject the null that all three are equal)

Another way to check for non-random assignment

- The ad assignment in this example was targeted towards customers who spent more in the past (who happened to be the youngest for this website). This is a common form of ad targeting in practice
- Another way to check for randomization would have been to regress treatment on past revenue per client (if you have this information in the data)

Q: Regress treatment on past revenue (measured in dollars). Interpret the slope coefficient. What about the intercept coefficient?

	coef	std err	t	P>|t|
Intercept	-0.4477	0.131	-3.406	0.001
past_rev	0.0842	0.014	6.023	0.000

A:

- This confirms the non-random assignment, since users with high past revenue are (stat. significantly) more likely to be treated (the probability of being treated increase by 8 percentage points for each dollar increase in past revenue)
- The intercept is negative. This is because the linear probability model can give you probabilities <0 and >1 so intercept is not always meaningful!

Omitted variable bias due to partial randomization

- **Q:** Ignore for now the non-random assignment and regress revenue on treatment. Do you think this regression over- or underestimates the effect of the ad?

	coef	std err	t	P>|t|
Intercept	8.8098	0.058	152.878	0.000
treatment	1.8281	0.099	18.470	0.000

- **A:** From previous slide treatment is positively correlated with past revenue. Past revenue is probably positively correlated with current revenue → by the omitted variable bias formula we have positive bias so regression overstates the effect of the add

- **Q:** Solution?
- **A:** The omitted variable is age (or past spending), so include the age bins as control variables

Making treatment random

- **Q:** Estimate a multivariate regression of revenue on treatment, controlling for the source of non-randomization (age bins). Compare the output to the univariate regression

A:

	coef	std err	t	P>|t|
Intercept	9.7054	0.120	80.955	0.000
treatment	1.2597	0.118	10.654	0.000
age_25to40	-1.1629	0.129	-9.033	0.000
age_above40	-0.8772	0.136	-6.469	0.000

- The treatment effect goes down after controlling for age
- Since treatment assignment is random once we control for age, 1.2597 is an unbiased (causal) estimate of the true causal effect of the ad on revenue

Same result by sequential estimation/1

- Let's verify that you would obtain the same result by the sequential estimation in the Frisch-Waugh theorem (this is just for intuition: in practice you just run the multivariate regression in the previous slide)
- Step 1: Regress treatment on age bin dummies and compute the residual from this regression. Then verify this residual is uncorrelated with age

```
#form residual from regression of treatment on age-bin dummies
result_1 = smf.ols(formula = 'treatment ~ age_25to40 + age_above40', data = targeted_ads).fit()
targeted_ads['residuals_1'] = result_1.resid

#treatment residual on age bin dummies
result_2 = smf.ols(formula = 'residuals_1 ~ age_25to40 + age_above40', data = targeted_ads).fit()
print(result_2.summary())
```

- We see below that the step 1 residual is uncorrelated with age, confirming that this residual isolates the random part of the treatment

	coef	std err	t	P>|t|
Intercept	2.166e-15	0.022	1e-13	1.000
age_25to40	-1.812e-15	0.030	-6.1e-14	1.000
age_above40	-6.922e-16	0.030	-2.31e-14	1.000

Same result by sequential estimation/2

- Step 2: Now regress revenue on the treatment residual from step 1

	coef	std err	t	P>|t|
Intercept	9.4295	0.052	180.994	0.000
residuals_1	1.2597	0.137	9.219	0.000

Q: How does this estimate relate to the estimate of the treatment effect from the multivariate regression of revenue on treatment and age bins?

A: They are the same due to the Frisch-Waugh theorem: the multivariate regression first isolates the random part of the treatment (the step 1 residual) and then estimates the effect on revenue of this random part only

	coef	std err	t	P>|t|
Intercept	9.7054	0.120	80.955	0.000
treatment	1.2597	0.118	10.654	0.000
age_25to40	-1.1629	0.129	-9.033	0.000
age_above40	-0.8772	0.136	-6.469	0.000

Should you also include “past revenue” as control?

Q: When estimating the ad effect, should you include “past revenue” as a control *in addition* to the age bins?

- Do you expect the treatment coefficient to change?
- Do you expect the standard error of the treatment effect to change?

A:

- It depends on whether the age bins eliminate all the non-random variation. You can check this by verifying that the step 1 residual (that eliminates non-randomness due to age) is uncorrelated with past revenue
- If so, the coefficient on treatment should not change much after also including past revenue because there is no further omitted variable bias in a regression that already controls for age
- The standard error shrinks when we add further controls, as we saw last week (as we saw this week, the intuition for this comes from the Frisch-Waugh theorem)

Is “past revenue” correlated with step 1 residual?

Q: Would you expect past revenue to be correlated with the residual from the step 1 regression? Regress the treatment residual from step 1 on past revenue to confirm your intuition.

	coef	std err	t	P>|t|
Intercept	0.0496	0.108	0.460	0.646
past_rev	-0.0053	0.011	-0.463	0.643

A: The step 1 regression isolates the random part of the treatment (the residual), which should thus be uncorrelated with all other variables such as past revenue. The fact that past_rev is not significant in the output confirms this

Including “past revenue” as additional control

A: Regression of revenue on treatment and age bins only

	coef	std err	t	P>|t|
Intercept	9.7054	0.120	80.955	0.000
treatment	1.2597	0.118	10.654	0.000
age_25to40	-1.1629	0.129	-9.033	0.000
age_above40	-0.8772	0.136	-6.469	0.000

Similar estimate of treatment effect but lower standard error when including past revenue as an additional control

	coef	std err	t	P>|t|
Intercept	-0.2403	0.341	-0.704	0.481
treatment	1.3003	0.086	15.201	0.000
age_25to40	-0.2229	0.098	-2.270	0.023
age_above40	-0.1716	0.101	-1.702	0.089
past_rev	1.0020	0.033	30.142	0.000

Re-cap: isolating random variation

- Sometimes we have experimental data where the treatment is partly randomly assigned → common for online ads that target users based on observable characteristics (age, gender, past purchases etc.)

- If we can measure these characteristics in the data, we can remove the endogeneity caused by the non-random assignment

- In the example, we know which part is random, which part is not:
  - Non-random assignment across age bins due to age-based targeting
  - Random assignment within age bins

- We can then remove the non-random part of the variation in treatment
  - Remove influence of non-random ad targeting by including the source of non-randomness (age or past revenue) as control in the regression
  - The Frisch-Waugh theorem explains why controlling for age isolates the random part of the treatment and eliminates the non-random part

Isolating random variation in observational data

- So far we have considered experimental data with partial randomization and used controls to isolate the random part of treatment assignment
- We can also use the same method in observational data if it contains random variation
- In observational data you typically try to find randomness that was accidental

Agenda today

- Econometric theory: sequential estimation of multivariate regression and “residual variation”
- Example: targeted advertising
- Workshop

Workshop: Estimating rideshare app demand function

- In the workshop you will address the important question of estimating a demand function (how quantity demanded reacts to price changes)
- Classically difficult problem due to endogeneity. We often say the problem is simultaneous causality – X causes Y and Y causes X (but we can also cast it as an omitted variable bias problem):
  - Prices and quantities are jointly determined by demand and supply

$$Demand : Q = \beta P + e$$
$$Supply : P = \gamma Q + u$$

  - In equilibrium

$$P = \frac{\gamma}{1 - \gamma\beta} e + \frac{1}{1 - \gamma\beta} u$$

so in the demand regression we have endogeneity because

$$Cov(P, e) = \frac{\gamma}{1 - \gamma\beta} Var(e) \neq 0$$

Workshop: Estimating rideshare app demand function

- There are different methods for eliminating the endogeneity due to simultaneous causality in demand estimation
- In the workshop, we will see two methods that are made possible by the nature of rideshared data and the pricing structure used by these apps (surge pricing)
    - Multivariate regression with control variables
    - Local regression