Lecture 9 - Economics 113

Professor Spearot

- Agenda
  1. Multivariate regression
  2. Simple - multiple regression comparison
  3. Bias

Multivariate Regression

- Example:

$$Grade = \beta_0 + \beta_1 hrs\_study + u$$

- We estimate $\widehat{\beta}_1 = 4.35$.
- Will a person who studies an extra hour per week get 4.35 more points?
  - Not if those who study more attend class more often.
  - *Attend* is an unobserved variable.
- Remember it must be that $E[u|x] = E[u|hrs\_study] = 0$

Multivariate Regression

- A new try:

$$Grade = \beta_0 + \beta_1 hrs\_study + \beta_2 Attend + u$$

  - *Attend* is number of classes attended.
- What is $E[u|x]$?
  - $E[u|hrs\_study, Attend] = 0$
  - If this holds, β_1 and β_2 will be unbiased estimates.
- Do you think that $E[u|hrs\_study, Attend] = 0$ is sensible?
- What else could be contained in u?

Multivariate Regression

Introduction

- Partial solution: Put in as many variables as possible.
- New estimating equation with k variables:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u$$

  - β_0 still an intercept
  - β_1, β_2 ,...β_k are slope parameters
  - Assume $E[u|x_1, x_2, \cdots, x_k] = 0$
- How do we estimate these models?
- Least squares techniques (derivatives!)
- Generate $\widehat{\beta}_1$, $\widehat{\beta}_2$,...$\widehat{\beta}_k$

Multivariate Regression

Using the estimates

- Predicted value $\widehat{y}$ is:

$$\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x_1 + \cdots + \widehat{\beta}_k x_k$$

- How do we interpret the $\widehat{\beta}' s$?
    - Holding everything constant (*ceteris paribus*), the effect of x_l is $\widehat{\beta}_l$.
- Example: Predicting who to admit to college.

Multivariate Regression

Using the estimates

- Suppose we estimated the following:

$$\widehat{FreshGPA} = 1.29 + 0.5HS\_GPA + 0.0003SAT$$

- What does 1.29 mean?
- Suppose two students have identical GPA's but SAT's of 1150 and 950.
- What is the predicted difference in *FreshGPA*?
- Take a difference:

$$\begin{aligned}\Delta\widehat{FreshGPA} &= 0.0003\Delta SAT \\ &= 0.0003(1150 - 950) = 0.6\end{aligned}$$

Multivariate Regression

Using the estimates

- Student A has $HS\_GPA = 3.2$, $SAT = 1250$.
- Student B has $HS\_GPA = 3.4$, $SAT = 1180$.
- To maximize $\widehat{FreshGPA}$, who do we admit?
- Generate Predictions, and Compare Them.
- Student A

$$\begin{aligned} \widehat{FreshGPA}_A &= 1.29 + 0.5 * 3.2 + 0.0003 * 1250 \\ &= 3.265 \end{aligned}$$

- Student B

$$\begin{aligned} \widehat{FreshGPA}_A &= 1.29 + 0.5 * 3.4 + 0.0003 * 1180 \\ &= 3.344 \end{aligned}$$

- We admit Student B!!

Multivariate Regression

Omitted variable bias

- Suppose that a *population* has the following relationship:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$

- However, we forget about x_2 and estimate:

$$\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x_1$$

- Our equation is *misspecified.*
- What is the relationship between the true β_1 and $\widehat{\beta}_1$?
- How do we quantify the bias?

Multivariate Regression

Omitted variable bias

- Population relationship between x_1 and x_2:

$$x_2 = \delta_0 + \delta_1 x_1 + \epsilon$$

- Plug x_2 into the population equation:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 \left(\delta_0 + \delta_1 x_1 + \epsilon\right) + u$$

- Expand:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 \delta_0 + \beta_2 \delta_1 x_1 + \beta_2 \epsilon + u$$

- Group terms:

$$y = \underbrace{\beta_0 + \beta_2 \delta_0}_{\widehat{\beta}_0} + \underbrace{(\beta_1 + \beta_2 \delta_1)}_{\widehat{\beta}_1} x_1 + \underbrace{\widetilde{u}}_{\beta_2 \epsilon + u}$$

Multivariate Regression

Omitted variable bias

- To quantify the bias, we need the sign of β_2 and δ_1.
- If x_2 and x_1 are positively correlated, $\delta_1 > 0$. If not, then $\delta_1 < 0$.
- If x_2 has a positive effect on y, $\beta_2 > 0$. If not, then $\beta_2 < 0$.
- Put them together:

	$Corr(x_1, x_2) > 0$ ($\delta_1 > 0$)	$Corr(x_1, x_2) < 0$ ($\delta_1 < 0$)
x_2 has a positive effect on y ($\beta_2 > 0$)	positive bias ($\beta_2\delta_1 > 0$)	negative bias ($\beta_2\delta_1 < 0$)
x_2 has a negative effect on y ($\beta_2 < 0$)	negative bias ($\beta_2\delta_1 < 0$)	positive bias ($\beta_2\delta_1 > 0$)

Multivariate Regression

Omitted variable bias - Examples

- Example: Effect of class attendance on grades
- Population follows:

$$final = \beta_0 + \beta_1 attend + \beta_2 study + u$$

- We estimate:

$$\widehat{final} = \widehat{\beta}_0 + \widehat{\beta}_1 attend$$

- Is $\widehat{\beta}_1$ over or under estimated?
- Holding *attend* constant, should *study* increase your grade?

  $\Longrightarrow$ Yes $\beta_2 > 0$.
- Are people that attend class more likely to study?

  $\Longrightarrow$ If so (probably true), then $\delta_1 > 0$.
- Bias in $\widehat{\beta}_1$?
- Bias is positive.

Multivariate Regression

Omitted variable bias - Examples

- Example: Effect of campaign spending on votes
- Population follows:

$$votes = \beta_0 + \beta_1 spend + \beta_2 popularity + u$$

- We estimate:

$$\widehat{votes} = \widehat{\beta}_0 + \widehat{\beta}_1 spend$$

- Is $\widehat{\beta}_1$ over or under estimated?
- Holding *popularity* constant, should *spend* increase *votes*?

$\Longrightarrow$ Yes $\beta_2 > 0$.

- Are more popular candidates likely to spend more?
  - If so, then $\delta_1 > 0 \Rightarrow \widehat{\beta}_1$ has a positive bias
    - Popularity $\Rightarrow$ Campaign Wealth $\Rightarrow$ Spend!!!
  - If not, then $\delta_1 < 0 \Rightarrow \widehat{\beta}_1$ has a negative bias
    - Spending to make-up for bad ideas

Multivariate Regression

Omitted variable bias - Examples

- Example: Effect of *GHG* on Temperature
- Population follows:

$$Temp = \beta_0 + \beta_1 GHG + \beta_2 Sun + u$$

- Sun is hard to measure. We estimate:

$$\widehat{Temp} = \widehat{\beta}_0 + \widehat{\beta}_1 GHG$$

- Is $\widehat{\beta}_1$ over or under estimated?
- Holding *GHG* constant, should *Sun* increase *Temp*?

  $\implies$ Yes: $\beta_2 > 0$.
- Is there a clear correlation between *Sun* and *GHC*?
  - Probably positive, but could differ depending on our sample.
  - Bias?
  - "Spurious" Correlation