Lecture 5: Multiple OLS Regression

Introduction to Econometrics,Fall 2018

Zhaopeng Qu

Nanjing University

10/11/2018

Review the last lecture

Simple OLS formula

- The linear regression model with one regressor is denoted by

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

- Where
  - Y_i is the **dependent variable**(Test Score)
  - X_i is the **independent variable** or regressor(Class Size or Student-Teacher Ratio)
  - u_i is the **error term** which contains all the other factors *besides* X that determine the value of the dependent variable, Y, for a specific observation, i.

The OLS Estimator

- The linear regression model with one regressor is denoted by

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

- Where
  - Y_i is the **dependent variable**(Test Score)
  - X_i is the **independent variable** or regressor(Class Size or Student-Teacher Ratio)
  - u_i is the **error term** which contains all the other factors *besides* X that determine the value of the dependent variable, Y, for a specific observation, i.

The OLS Estimator

- The estimators of the slope and intercept that *minimize the sum of the squares* of $\hat{u}_i$,thus

$$arg\,\min_{b_0,b_1} \sum_{i=1}^{n} \hat{u}_i^2 = \min_{b_0,b_1} \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i)^2$$

are called the **ordinary least squares (OLS) estimators** of β_0 and β_1.

OLS estimator of β_1:

$$b_1 = \hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})(X_i - \overline{X})}$$

Least Squares Assumptions

Under 3 least squares assumptions,

1. Assumption 1
2. Assumption 2
3. Assumption 3

the OLS estimators will be

- **unbiased**
- **consistent**
- **normal sampling distribution**

Simple OLS Regression v.s. RCT

- Regression is a way to control observable confounding factors, Which assume the source of selection bias is only from the difference in observed characteristics.
- In a simple regression model, OLS estimators are just a generalizing continuous version of RCT when least squares assumptions are hold.
- But in contrast to RCT, in observational studies, researchers cannot control the assignment of treatment into a treatment group versus a control group.
- To make two groups comparable, we need to keep treatment and control group "**other thing equal**"in observed characteristics and unobserved characteristics.

Multiple OLS Regression: Introduction

Violation of the first Least Squares Assumption

- Recall simple OLS regression equation

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

- **Question: What does u_i represent?**
  - Answer: contains all other factors(variables) which potentially affect Y_i.
- **Assumption 1**

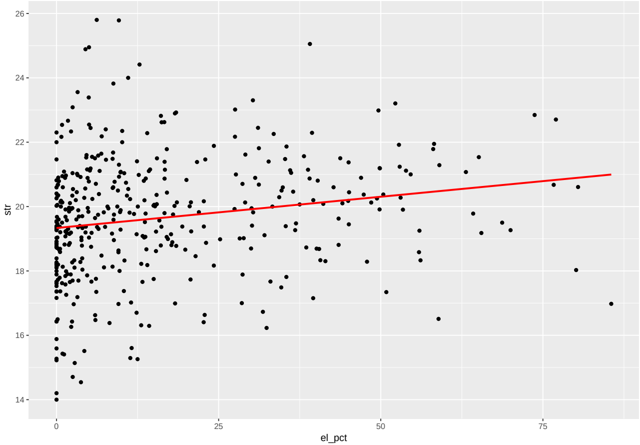
$$E(u_i|X_i) = 0$$

  - It states that u_i are unrelated to X_i in the sense that,given a value of X_i,the mean of these other factors equals **zero**.
  - But what if they (or at least one) are *correlated* with X_i?

Example: Class Size and Test Score

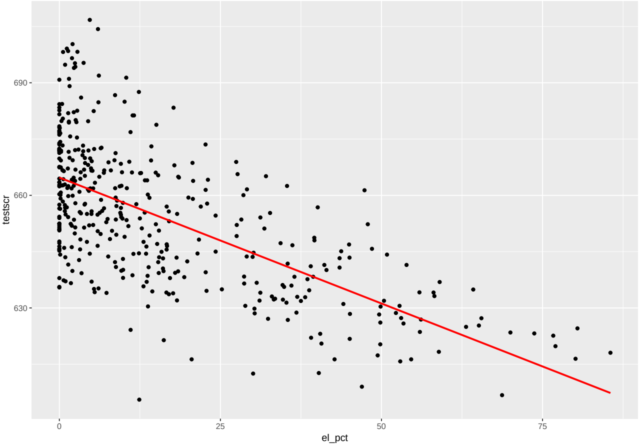
- Many other factors can affect student's performance in the school.
- One of other factors is the share of immigrants in the class(school,district). Because immigrant children may have different backgrouds from native children, such as
    - parents'education level
    - family income and wealth
    - preparenting style
    - traditonal culture

Scatter Plot: english learners and STR



- higher share of english learner, bigger class size

Scatter Plot: english learners and testscr



- higher share of english learner, lower testscore

English learner as an Omitted Variable

- Class size may be related to percentage of English learners and students who are still learning English likely have lower test scores.
- It implies that percentage of English learners is contained in u_i, in turn that Assumption 1 is violated.
- It means that the estimates of $\hat{\beta}_1$ and $\hat{\beta}_0$ are *biased* and *inconsistent*.

English learner as an Omitted Variable

- As before, X_i and Y_i represent STR and Test Score.
- Besides, W_i is the variable which represents the share of English learners.
- Suppose that we have no information about it for some reasons, then we have to omit in the regression.
- Then we have two regression:
  - **True model**(Long regression):

$$Y_i = \beta_0 + \beta_1 X_i + \gamma W_i + u_i$$

    where $E(u_i|X_i, W_i) = 0$
  - **OVB model**(Short regression):

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

    where $v_i = \gamma W_i + u_i$

Omitted Variable Bias: Biasedness

- Let us see what is the consequece of OVB

$$E[\hat{\beta}_1] = E\left[\frac{\sum(X_i - \bar{X})(\beta_0 + \beta_1 X_i + v_i - (\beta_0 + \beta_1 \overline{X} + \overline{v}))}{\sum(X_i - \bar{X})(X_i - \bar{X})}\right]$$
$$= E\left[\frac{\sum(X_i - \bar{X})(\beta_0 + \beta_1 X_i + \gamma W_i + u_i - (\beta_0 + \beta_1 \overline{X} + \gamma \overline{W} + \overline{u}))}{\sum(X_i - \bar{X})(X_i - \bar{X})}\right]$$

- Skip Several steps in algebra which is very **similar** to procedures for proving unbiasedness of β
- At last, we get (**Please prove it by yourself**)

$$E[\hat{\beta}_1] = \beta_1 + \gamma E\left[\frac{\sum(X_i - \bar{X})(W_i - \bar{W})}{\sum(X_i - \bar{X})(X_i - \bar{X})}\right]$$

Omitted Variable Bias: Biasedness

- As proving unbiasedness of $\hat{\beta}_1$,we can know
- If W_i is unrelated to X_i,then $E[\hat{\beta}_1] = \beta_1$, because

$$E\left[\frac{\sum(X_i - \bar{X})(W_i - \bar{W})}{\sum(X_i - \bar{X})(X_i - \bar{X})}\right] = 0$$

- If W_i is no determinant of Y_i, which means that

$$\gamma = 0$$

, then $E[\hat{\beta}_1] = \beta_1$, too,

- Only if **both two conditions** above are violated *simultaneously*, then $\hat{\beta}_1$ is **biased**, which is normally called **Omitted Variable Bias**.

Omitted Variable Bias(OVB): inconsistency

- Recall: consistency when n is large, thus
- OLS with on OVB

$$plim\hat{\beta}_1 = \frac{Cov(X_i, Y_i)}{Var(X_i)}$$

Omitted Variable Bias(OVB): inconsistency

$$
\begin{aligned}
plim\hat{\beta}_1 &= \frac{Cov(X_i, Y_i)}{VarX_i} \\
&= \frac{Cov(X_i, (\beta_0 + \beta_1 X_i + v_i))}{VarX_i}
\end{aligned}
$$

Omitted Variable Bias(OVB): inconsistency

$$
\begin{aligned}
plim\hat{\beta}_1 &= \frac{Cov(X_i, Y_i)}{VarX_i} \\
&= \frac{Cov(X_i, (\beta_0 + \beta_1 X_i + v_i))}{VarX_i} \\
&= \frac{Cov(X_i, (\beta_0 + \beta_1 X_i + \gamma W_i + u_i))}{VarX_i}
\end{aligned}
$$

Omitted Variable Bias(OVB): inconsistency

$$
\begin{aligned}
plim\hat{\beta}_1 &= \frac{Cov(X_i, Y_i)}{VarX_i} \\
&= \frac{Cov(X_i, (\beta_0 + \beta_1 X_i + v_i))}{VarX_i} \\
&= \frac{Cov(X_i, (\beta_0 + \beta_1 X_i + \gamma W_i + u_i))}{VarX_i} \\
&= \frac{Cov(X_i, \beta_0) + \beta_1 Cov(X_i, X_i) + \gamma Cov(X_i, W_i) + Cov(X_i, u_i)}{VarX_i}
\end{aligned}
$$

Omitted Variable Bias(OVB): inconsistency

$$
\begin{aligned}
plim\hat{\beta}_1 &= \frac{Cov(X_i, Y_i)}{VarX_i} \\
&= \frac{Cov(X_i, (\beta_0 + \beta_1 X_i + v_i))}{VarX_i} \\
&= \frac{Cov(X_i, (\beta_0 + \beta_1 X_i + \gamma W_i + u_i))}{VarX_i} \\
&= \frac{Cov(X_i, \beta_0) + \beta_1 Cov(X_i, X_i) + \gamma Cov(X_i, W_i) + Cov(X_i, u_i)}{VarX_i} \\
&= \beta_1 + \gamma \frac{Cov(X_i, W_i)}{VarX_i}
\end{aligned}
$$

Omitted Variable Bias(OVB): inconsistency

- Thus we obtain

$$plim\hat{\beta}_1 = \beta_1 + \gamma\frac{Cov(X_i, W_i)}{VarX_i}$$

- $\hat{\beta}_1$ is still consistent
  - if W_i is unrelated to X, thus $Cov(X_i, W_i) = 0$
  - if W_i has no effect on Y_i, thus $\gamma = 0$
- if both two conditions above hold *simultaneously*, then $\hat{\beta}_1$ is **inconsistent**.

Omitted Variable Bias(OVB):Directions

- If OVB can be possible in our regression,then we should guess the **directions** of the bias, in case that we can't eliminate it.
- Summary of the bias when w_i is omitted in estimating equation

	$Cov(X_i, W_i) > 0$	$Cov(X_i, W_i) < 0$
$\gamma > 0$	Positive bias	Negative bias
$\gamma < 0$	Negative bias	Positive bias

Omiited Variable Bias: Examples

- Question: If we omit following variables, then what are the directions of these biases? and why?
  1. Time of day of the test
  2. Parking lot space per pupil
  3. Teachers' Salary
  4. Family income
  5. Percentage of English learners

Omiited Variable Bias: Examples

- Regress *Testscore* on *Class size*

```
##
## Call:
## lm(formula = testscr ~ str, data = ca)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -47.727 -14.251   0.483  12.822  48.540
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 698.9330     9.4675  73.825  < 2e-16 ***
## str          -2.2798     0.4798  -4.751 2.78e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '
##
```

Omiited Variable Bias: Examples

- Regress *Testscore* on *Class size* and *the percentage of English learners*

```
##
## Call:
## lm(formula = testscr ~ str + el_pct, data = ca)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -48.845 -10.240  -0.308   9.815  43.461
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 686.03225    7.41131  92.566  < 2e-16 ***
## str          -1.10130    0.38028  -2.896  0.00398 **
## el_pct       -0.64978    0.03934 -16.516  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '
```

Omiited Variable Bias: Examples

表 **2:** Class Size and Test Score

	Dependent variable:	
	testscr	
	(1)	(2)
str	-2.280^{***}	-1.101^{***}
	(0.480)	(0.380)
el_pct		-0.650^{***}
		(0.039)
Constant	698.933^{***}	686.032^{***}
	(9.467)	(7.411)
Observations	420	420
R^2	0.051	0.426
Note:	*p<0.1; **p<0.05; ***p<0.01	

Warp Up

- OVB bias is the most possible bias when we run OLS regression using nonexperiemental data.
- The simplest way to overcome OVB: **control it**.

Multiple OLS Regression: Estimation

Multiple regression model with k regressors

- The multiple regression model is

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + ... + \beta_k X_{k,i} + u_i, i = 1, ..., n$$

where

- Y_i is the *dependent variable*
- $X_1, X_2, ...X_k$ are the *independent variables(includes some control variables)*
- $\beta_i, j = 1...k$ are slope coefficients on X_i corresponding.
- β_0 is the estimate *intercept*, the value of Y when all $X_j = 0, j = 1...k$
- u_i is the regression error term.

Interpretation of coefficients

- β_j is partial (marginal) effect of X_j on Y.

$$\beta_j = \frac{\partial Y_i}{\partial X_{j,i}}$$

- β_j is also partial (marginal) effect of $E[Y_i|X_1..X_k]$.

$$\beta_j = \frac{\partial E[Y_i|X_1, ..., X_k]}{\partial X_{j,i}}$$

- it does mean "other things equal", thus the concept of **ceteris paribus**

Independent Variable v.s Control Variables

- Generally, we would like to pay more attention to **only one** independent variable(thus we would like to call it **treatment variable**), though there could be many independent variables.
- Other variables in the right hand of equation, we call them **control variables**, which we would like to explicitly hold fixed when studying the effect of X_1 on Y.
- More specifically,regression model turns into

$$Y_i = \beta_0 + \beta_1 D_i + \gamma_2 C_{2,i} + ... + \gamma_k C_{k,i} + u_i, i = 1, ..., n$$

- transform it into

$$Y_i = \beta_0 + \beta_1 D_i + C_{2...k,i}\gamma'_{2...k} + u_i, i = 1, ..., n$$

OLS Estimation in Multiple Regressors

- As in simple OLS, the estimator multiple Regression is just a minimize the following question

$$argmin \sum_{b_0,b_1,...,b_k} (Y_i - b_0 - b_1 X_{1,i} - ... - b_k X_{k,i})^2$$

OLS Estimation in Multiple Regressors

- The OLS estimators $\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_k$ are obtained by solving the following **system of normal equations**

$$\begin{aligned}
\sum \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1,i} - ... - \hat{\beta}_k X_{k,i}\right) &= 0 \\
\sum \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1,i} - ... - \hat{\beta}_k X_{k,i}\right) X_{1,i} &= 0 \\
\vdots &= \vdots \\
\sum \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1,i} - ... - \hat{\beta}_k X_{k,i}\right) X_{k,i} &= 0
\end{aligned}$$

OLS Estimation in Multiple Regressors

- Since the fitted residuals are

$$\hat{u}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1,i} - ... - \hat{\beta}_k X_{k,i}$$

- the normal equations can be written as

$$\begin{aligned} \sum \hat{u}_i &= 0 \\ \sum \hat{u}_i X_{1,i} &= 0 \\ \vdots &= \vdots \\ \sum \hat{u}_i X_{k,i} &= 0 \end{aligned}$$

Partitioned regression: OLS estimators in Multiple Regression

Introduction

If the four least squares assumptions in the multiple regression model hold:

- The OLS estimators $\hat{\beta}_0, \hat{\beta}_1 ... \hat{\beta}_k$ are unbiased.
- The OLS estimators $\hat{\beta}_0, \hat{\beta}_1 ... \hat{\beta}_k$ are consistent.
- The OLS estimators $\hat{\beta}_0, \hat{\beta}_1 ... \hat{\beta}_k$ are normally distributed in large samples.
- Formal proofs need to use the knowledge of **linear algebra**, thus **the matrix**. We only prove them in a simple case.

Partitioned regression: OLS estimators

- A useful representation of $\hat{\beta}_j$ could be obtained by the **partitioned regression**.
- Suppose we want to obtain an expression for $\hat{\beta}_1$.
- Regress $X_{1,i}$ on other regressors,thus

Partitioned regression: OLS estimators

- $$X_{1,i} = \hat{\gamma}_0 + \hat{\gamma}_2 X_{2,i} + ... + \hat{\gamma}_k X_{k,i} + \tilde{X}_{1,i}$$

  where $\tilde{X}_{1,i}$ is the fitted OLS residual(just a variation of u_i)
- Then we could prove that

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} \tilde{X}_{1,i} Y_i}{\sum_{i=1}^{n} \tilde{X}_{1,i}^2}$$

Proof of Partitioned regression result(1)

- we know

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1,i} + \hat{\beta}_2 X_{2,i} + ... + \hat{\beta}_k X_{k,i} + \hat{u}_i$$

where $\sum \hat{u}_i = \sum \hat{u}_i X_{ji} = 0, j = 1, 2, ..., k$

- Now

$$\begin{aligned}
\frac{\sum_{i=1}^n \tilde{X}_{1,i} Y_i}{\sum_{i=1}^n \tilde{X}_{1,i}^2} &= \frac{\sum \tilde{X}_{1,i}(\hat{\beta}_0 + \hat{\beta}_1 X_{1,i} + \hat{\beta}_2 X_{2,i} + ... + \hat{\beta}_k X_{k,i} + \hat{u}_i)}{\sum \tilde{X}_{1,i}^2} \\
&= \hat{\beta}_0 \frac{\sum_{i=1}^n \tilde{X}_{1,i}}{\sum_{i=1}^n \tilde{X}_{1,i}^2} + \hat{\beta}_1 \frac{\sum_{i=1}^n \tilde{X}_{1,i} X_{1,i}}{\sum_{i=1}^n \tilde{X}_{1,i}^2} + ... \\
&+ \hat{\beta}_k \frac{\sum_{i=1}^n \tilde{X}_{1,i} X_{k,i}}{\sum_{i=1}^n \tilde{X}_{1,i}^2} + \frac{\sum_{i=1}^n \tilde{X}_{1,i} \hat{u}_i}{\sum_{i=1}^n \tilde{X}_{1,i}^2}
\end{aligned}$$

Proof of Partitioned regression result(2)

- $\tilde{X}_{1,i}$ is the fitted OLS residual for the regression

$$X_{1,i} = \hat{\gamma}_0 + \hat{\gamma}_2 X_{2,i} + ... + \hat{\gamma}_k X_{k,i} + \tilde{X}_{1,i}$$

- so it is a variation of $\hat{u}_i$, then we have

$$\sum_{i=1}^{n} \tilde{X}_{1,i} = 0 \; and \; \sum_{i=1}^{n} \tilde{X}_{1,i} X_{j,i} = 0 \; , j = 2, 3, ..., k$$

Proof of Partitioned regression result(3)

- We also have

$$
\begin{aligned}
&\sum_{i=1}^{n} \tilde{X}_{1,i} X_{1,i} \\
&= \sum_{i=1}^{n} \tilde{X}_{1,i} (\hat{\gamma}_0 + \hat{\gamma}_2 X_{2,i} + ... + \hat{\gamma}_k X_{k,i} + \tilde{X}_{1,i}) \\
&= \hat{\gamma}_0 \cdot 0 + \hat{\gamma}_2 \cdot 0 + ... + \hat{\gamma}_k \cdot 0 + \sum \tilde{X}_{1,i}^2 \\
&= \sum \tilde{X}_{1,i}^2
\end{aligned}
$$

Proof of Partitioned regression result(4)

- Recall: $\hat{u}_i$ are the fitted residuals from the regression of Y against all X, then

$$\sum_{i=1}^{n} \hat{u}_i = \sum_{i=1}^{n} \hat{u}_i X_{j,i} = 0 \ , j = 1, 2, 3, ..., k$$

Proof of Partitioned regression result(5)

- Recall: $\hat{u}_i$ are the fitted residuals from the regression of Y against all X, then

$$\sum_{i=1}^{n}\hat{u}_i=\sum_{i=1}^{n}\hat{u}_iX_{j,i}=0\ ,j=1,2,3,...,k$$

- We also have

$$\begin{aligned}&\sum_{i=1}^{n}\tilde{X}_{1,i}\hat{u}_i\\&=\sum_{i=1}^{n}(X_{1,i}-\hat{\gamma}_0-\hat{\gamma}_2X_{2,i}-...-\hat{\gamma}_kX_{k,i})\hat{u}_i\\&=0-\hat{\gamma}_0\cdot 0-\hat{\gamma}_2\cdot 0-...-\hat{\gamma}_k\cdot 0\\&=0\end{aligned}$$

wrap up so far

- OLS Regression
- and $\tilde{X}_{1,i}$ is the fitted OLS residual for the regression

$$X_{1,i} = \hat{\gamma}_0 + \hat{\gamma}_2 X_{2,i} + ... + \hat{\gamma}_k X_{k,i} + \tilde{X}_{1,i}$$

- we obtained

$$\sum_{i=1}^{n} \tilde{X}_{1,i} = \sum_{i=1}^{n} \tilde{X}_{1,i} X_{j,i} = 0 \ , j = 2, 3, ..., k$$

$$\sum_{i=1}^{n} \tilde{X}_{1,i} X_{1,i} = \sum \tilde{X}_{1,i}^2$$

$$\sum_{i=1}^{n} \tilde{X}_{1,i} \hat{u}_i = 0$$

Proof of Partitioned regression result(6)

- we have shown that

$$\frac{\sum_{i=1}^{n}\tilde{X}_{1,i}Y_i}{\sum_{i=1}^{n}\tilde{X}_{1,i}^2}=\hat{\beta}_0\frac{\sum_{i=1}^{n}\tilde{X}_{1,i}}{\sum_{i=1}^{n}\tilde{X}_{1,i}^2}+\hat{\beta}_1\frac{\sum_{i=1}^{n}\tilde{X}_{1,i}X_{1,i}}{\sum_{i=1}^{n}\tilde{X}_{1,i}^2}+...$$
$$+\hat{\beta}_k\frac{\sum_{i=1}^{n}\tilde{X}_{1,i}X_{k,i}}{\sum_{i=1}^{n}\tilde{X}_{1,i}^2}+\frac{\sum_{i=1}^{n}\tilde{X}_{1,i}\hat{u}_i}{\sum_{i=1}^{n}\tilde{X}_{1,i}^2}$$

- then

$$\frac{\sum_{i=1}^{n}\tilde{X}_{1,i}Y_i}{\sum_{i=1}^{n}\tilde{X}_{1,i}^2}=\hat{\beta}_1$$

.

- Identical argument works for $j=2,3,...,k$, thus

$$\hat{\beta}_j=\frac{\sum_{i=1}^{n}\tilde{X}_{j,i}Y_i}{\sum_{i=1}^{n}\tilde{X}_{j,i}^2}$$

The intuition of Partitioned regression

Partialling Out

- First, we regress X_j against the rest of the regressors (and a constant) and keep $\tilde{X}_j$ which is the "part" of X_j that is **uncorrelated**
- Then, to obtain $\hat{\beta}_j$, we regress Y against $\tilde{X}_j$ which is *"clean"* from correlation with other regressors.
- $\hat{\beta}_j$ measures the effect of X_1 after the effects of $X_2, ..., X_k$ have been *partialled out or netted out*.

Example: Test scores and Student Teacher Ratios

```
tilde.str <- residuals(lm(str ~ el_pct+avginc, data=ca))
mean(tilde.str) # should be zero
```

```
## [1] 1.305121e-17
```

```
sum(tilde.str)  # also is zero
```

```
## [1] 5.412337e-15
```

```
cov(tilde.str,ca$avginc)# should be zero too
```

```
## [1] 3.650126e-16
```

Example: Test scores and Student Teacher Ratios(2)

```
tilde.str_str <- tilde.str*ca$str # uX
tilde.strstr <- tilde.str^2
sum(tilde.str_str) # sum(uX)=sum(u^2)
```

```
## [1] 1396.348
```

```
sum(tilde.strstr)# should be equal the result above.
```

```
## [1] 1396.348
```

Example: Test scores and Student Teacher Ratios(3)

```
sum(tilde.str*ca$testscr)/sum(tilde.str^2)
```

```
## [1] -0.06877552
```

Example: Test scores and Student Teacher Ratios(3)

```
##
## Call:
## lm(formula = testscr ~ tilde.str, data = ca)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -48.50 -14.16   0.39  12.57  52.57
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 654.15655    0.93080 702.790   <2e-16 ***
## tilde.str    -0.06878    0.51049  -0.135    0.893
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '
##
## Residual standard error: 19.08 on 418 degrees of freedom
```

Example: Test scores and Student Teacher Ratios(4)

```
reg4 <- lm(testscr ~ str+el_pct+avginc,data = ca)
summary(reg4)
```

```
##
## Call:
## lm(formula = testscr ~ str + el_pct + avginc, data = ca)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -42.800  -6.862   0.275   6.586  31.199
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 640.31550    5.77489 110.879   <2e-16 ***
## str          -0.06878    0.27691  -0.248    0.804
## el_pct       -0.48827    0.02928 -16.674   <2e-16 ***
```

Measures of Fit in Multiple Regression

Standard Error of the Regression

- Recall: SER(Standard Error of the Regression)
  - SER is an **estimator** of the standard deviation of the u_i, which are measures of the spread of the Y's around the regression line.
  - Because the regression errors are unobserved, the SER is computed using their sample counterparts, the OLS residuals $\hat{u}_i$

$$SER = s_{\hat{u}} = \sqrt{s_{\hat{u}}^2}$$

where $s_{\hat{u}}^2 = \frac{1}{n-k-1}\sum \widehat{u^2}_i = \frac{SSR}{n-k-1}$

- $n-k-1$ because we have $k+1$ stricted conditions in the F.O.C.In another word,in order to construct $\widehat{u^2}_i$, we have to estimate $k+1$ parameters,thus $\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_k$

Measures of Fit in Multiple Regression

- Actual = Predicted+residual: $Y_i = \hat{Y}_i + \hat{u}_i$
- The regression R^2 is the fraction of the sample variance of Y_i explained by (or predicted by) the regressors.

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{SSR}{TSS}$$

- R^2 *always increases when you add another regressor*. Because in general the SSR will decrease.

Measures of Fit: The Adjusted R^2

- the adjusted R^2,is a modified version of the R^2 that does not necessarily increase when a new regressor is added.

$$\overline{R^2} = 1 - \frac{n-1}{n-k-1}\frac{SSR}{TSS} = 1 - \frac{s_{\hat{u}}^2}{s_Y^2}$$

- because $\frac{n-1}{n-k-1}$ is always greater than 1, so $\overline{R^2} < R^2$
- adding a regressor has two opposite effects on the $\overline{R^2}$.
- $\overline{R^2}$ can be negative.
- **Remind**: *neither R^2 nor $\overline{R^2}$ is not the golden criterion for good or bad OLS estimation.*

Multiple regression: Assumpition

Multiple regression: Assumpition

- Assumption 1: The conditional distribution of u_i given $X_{1i},...,X_{ki}$ has mean zero,thus

$$E[u_i|X_{1i},...,X_{ki}] = 0$$

- Assumption 2: $(Y_i, X_{1i},...,X_{ki})$ are i.i.d.
- Assumption 3: Large outliers are unlikely.
- Assumption 4: No perfect multicollinearity.

Perfect multicollinearity

Perfect multicollinearity arises when one of the regressors is a **perfect** linear combination of the other regressors.

- Binary variables are sometimes referred to as dummy variables
- If you include a full set of binary variables (a complete and mutually exclusive categorization) and an intercept in the regression, you will have perfect multicollinearity.
  - eg. female and male = 1-female
  - eg. West, Central and East China
- This is called the **dummy variable trap**.
- Solutions to the dummy variable trap: Omit one of the groups or the intercept

Perfect multicollinearity

- regress *Testscore* on *Class size* and *the percentage of English learners*

```
##
## Call:
## lm(formula = testscr ~ str + el_pct, data = ca)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -48.845 -10.240  -0.308   9.815  43.461
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 686.03225    7.41131  92.566  < 2e-16 ***
## str          -1.10130    0.38028  -2.896  0.00398 **
## el_pct       -0.64978    0.03934 -16.516  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '
```

Perfect multicollinearity

- add a new variable nel=1-el_pct into the regression

```
##
## Call:
## lm(formula = testscr ~ str + nel_pct + el_pct, data = ca)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -48.845 -10.240  -0.308   9.815  43.461
##
## Coefficients: (1 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 685.38247    7.41556  92.425  < 2e-16 ***
## str          -1.10130    0.38028  -2.896  0.00398 **
## nel_pct       0.64978    0.03934  16.516  < 2e-16 ***
## el_pct             NA         NA      NA       NA
## ---
```

Perfect multicollinearity

表 **3:** Class Size and Test Score

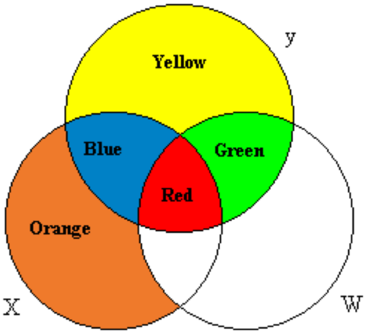
	Dependent variable:	
	testscr	
	(1)	(2)
str	-1.101^{***}	-1.101^{***}
	(0.380)	(0.380)
nel_pct		0.650^{***}
		(0.039)
el_pct	-0.650^{***}	
	(0.039)	
Constant	686.032^{***}	685.382^{***}
	(7.411)	(7.416)
Observations	420	420
R^2	0.426	0.426
Note:	*p<0.1; **p<0.05; ***p<0.01	

Multicollinearity

Multicollinearity means that two or more regressors are **highly** correlated, but one regressor is **NOT** a perfect linear function of one or more of the other regressors.

- **multicollinearity** is **NOT** a violation of OLS assumptions.
  - It does not impose theoretical problem for the calculation of OLS estimators.
- But if two regressors are highly correlated, then the the coefficient on at least one of the regressors is imprecisely estimated (high variance).
- to what extent two correlated variables can be seen as "highly correlated"?
  - **rule of thumb**: correlation coefficient is over **0.8**.

Venn Diagrams for Multiple Regression Model



1) In a simple model (y on X), OLS uses Blue + Red to estimate β. 2) When y is regressed on X and W: OLS throws away the red area and just uses blue to estimate β. 3) Idea: red area is contaminated(we do not know if the movements in y are due to X or to W).

Venn Diagrams for Multicollinearity

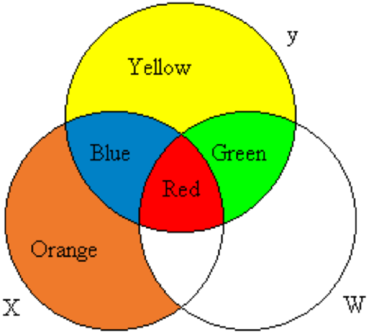


Figure 3a Modest collinearity

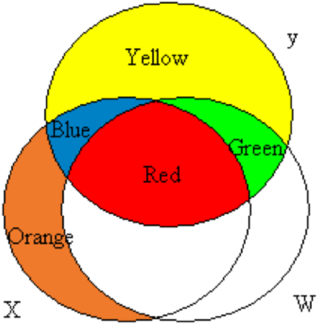


Figure 3b Considerable collinearity

- less information (compare the Blue and Green areas in both figures) is used, the estimation is less precise.

Multiple regression model: class size example

表 **4:** Class Size and Test Score

	Dependent variable:		
	testscr		
	(1)	(2)	(3)
str	-2.280^{***}	-1.101^{***}	-0.069
	(0.480)	(0.380)	(0.277)
el_pct		-0.650^{***}	-0.488^{***}
		(0.039)	(0.029)
avginc			1.495^{***}
			(0.075)
Constant	698.933^{***}	686.032^{***}	640.315^{***}
	(9.467)	(7.411)	(5.775)
Observations	420	420	420
R^2	0.051	0.426	0.707
Adjusted R^2	0.049	0.424	0.705

Properties of OLS estimator in Multiple Regression

Proof that OLS is unbiased(1)

- Use partitioned regression formula

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} \tilde{X}_{1,i} Y_i}{\sum_{i=1}^{n} \tilde{X}_{1,i}^2}$$

- Substitute

  $Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + ... + \beta_k X_{k,i} + u_i, i = 1, ..., n$,then

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\sum \tilde{X}_{1,i}(\beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + ... + \beta_k X_{k,i} + u_i)}{\sum \tilde{X}_{1,i}^2} \\
&= \beta_0 \frac{\sum_{i=1}^{n} \tilde{X}_{1,i}}{\sum_{i=1}^{n} \tilde{X}_{1,i}^2} + \beta_1 \frac{\sum_{i=1}^{n} \tilde{X}_{1,i} X_{1,i}}{\sum_{i=1}^{n} \tilde{X}_{1,i}^2} + ... \\
&+ \beta_k \frac{\sum_{i=1}^{n} \tilde{X}_{1,i} X_{k,i}}{\sum_{i=1}^{n} \tilde{X}_{1,i}^2} + \frac{\sum_{i=1}^{n} \tilde{X}_{1,i} u_i}{\sum_{i=1}^{n} \tilde{X}_{1,i}^2}
\end{aligned}$$

Proof that OLS is unbiased(2)

- Because

$$\sum_{i=1}^{n}\tilde{X}_{1,i}=\sum_{i=1}^{n}\tilde{X}_{1,i}X_{j,i}=0\ ,j=2,3,...,k$$

$$\sum_{i=1}^{n}\tilde{X}_{1,i}X_{1,i}=\sum\tilde{X}_{1,i}^{2}$$

- Therefore

$$\hat{\beta}_1=\beta_1+\frac{\sum_{i=1}^{n}\tilde{X}_{1,i}u_i}{\sum_{i=1}^{n}\tilde{X}_{1,i}^{2}}$$

Proof that OLS is unbiased(3)

- we have that

$$\hat{\beta}_1 = \beta_1 + \frac{\sum \tilde{X}_{1,i} u_i}{\sum \tilde{X}_{1,i}^2}$$

- Take expectations of $\hat{\beta}_1$ and based on **Assumption 1** again

$$\begin{aligned} E[\hat{\beta}_1] &= E\Big[E[\hat{\beta}_1|X]\Big] \\ &= \beta_1 + 0 \end{aligned}$$

- Identical argument works for $j = 2, 3, ..., k$

The Distribution of the OLS Estimators

- In addition, in large samples, the sampling distribution of $\hat{\beta}_1$ and $\hat{\beta}_0$ is well approximated by a bivariate normal distribution.
- Under the least squares assumptions,the OLS estimators $\hat{\beta}_1$ and $\hat{\beta}_0$, are unbiased and consistent estimators of
- The OLS estimators are averages of the randomly sampled data, and if the sample size is sufficiently large, the sampling distribution of those averages becomes normal. Because the multivariate normal distribution is best handled mathematically using matrix algebra, the expressions for the joint distribution of the OLS estimators are deferred to Chapter 18(SW textbook).
- If the least squares assumptions hold, then in large samples the OLS estimators $\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_k$ are jointly normally distributed and each

$$\hat{\beta}_j \sim N(\beta_j, \sigma^2_{\hat{\beta}_j})\ , j = 0, ..., k$$