Simple and Multiple OLS Regression

Introduction to Econometrics,Fall 2017

Zhaopeng Qu

Nanjing University

10/16/2017

Review the last lecture

CEF(conditional expectation function)

- CEF is a natural summary of the relationship between Y and X. If we can know CEF, then we can describe the relationship of Y and X.
- Regression estimates provides a valuable baseline for almost all empirical research because Regression is tightly linked to CEF
- if CEF is linear, then OLS regression is it.
- if CEF is nonlinear, then OLS regression provides a best linear approximation to it under MMSE condition.

OLS with One Regressor: Estimation

The OLS estimators

- Question of interest: What is the effect of a change in X_i(Class Size) on Y_i(Test Score)

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

- Last week we derived the OLS estimators of β_0 andβ_1:

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

$$\hat{\beta}_1 = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})(X_i - \bar{X})}$$

Least Squares Assumptions

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

if the 3 least squares assumptions hold the OLS estimators

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

Properties of the OLS estimator: unbiasedness

- take expectation to β_0 :

$$E[\hat{\beta}_0] = \bar{Y} - E[\hat{\beta}_1]\bar{X}$$

- if β_1 is unbiased, then β_0 is also unbiased.
- Remind we have

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

$$\overline{Y} = \beta_0 + \beta_1 \overline{X} + \overline{u}$$

- So take expectation to β_1:

$$E[\hat{\beta}_1] = E\left[\frac{\sum(X_i - \bar{X})/(Y_i - \bar{Y})}{\sum(X_i - \bar{X})(X_i - \bar{X})}\right]$$

Properties of the OLS estimator: unbiasedness

- Continued

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

$$= E\left[\frac{\sum(X_i - \bar{X})(\beta_1(X_i - \overline{X}) + (u_i - \overline{u}))}{\sum(X_i - \bar{X})(X_i - \bar{X})}\right]$$

$$= \beta_1 + E\left[\frac{\sum(X_i - \bar{X})(u_i - \overline{u})}{\sum(X_i - \bar{X})(X_i - \bar{X})}\right]$$

- Because $\sum \overline{u} = 0$ and $\sum \overline{u} X_i = 0$, so

$$= \beta_1 + E\left[\frac{\sum(X_i - \overline{X})u_i}{\sum(X_i - \overline{X})(X_i - \overline{X})}\right]$$

Properties of the OLS estimator: unbiasedness

- Continued

$$= \beta_1 + E\left[\frac{\sum(X_i - \overline{X})u_i}{\sum(X_i - \overline{X})(X_i - \overline{X})}\right]$$

- then then we could obtain

$$E[\hat{\beta}_1] = \beta_1 \; if \; E[u_i|X_i] = 0$$

- thus both β_0 and β_1 are **unbiased** on the condition of **Assumption 1**.

Properties of the OLS estimator: Consistency

- *Notation*: $\hat{\beta}_1 \xrightarrow{p} \beta_1$ or $plim\hat{\beta}_1 = \beta_1$, so

$$plim\hat{\beta}_1 = plim\Big[\frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})(X_i - \bar{X})}\Big]$$

$$plim\hat{\beta}_1 = plim\Big[\frac{\frac{1}{n-1}\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\frac{1}{n-1}\sum(X_i - \bar{X})(X_i - \bar{X})}\Big] = plim\Big(\frac{s_{xy}}{s_x^2}\Big)$$

where s_{xy} and s_x^2 are sample covariance and sample variance.

Properties of the OLS estimator: Consistency

- *Continuous Mapping Theorem*: For every continuous function $g(t)$ and random variable X:

$$plim(g(X)) = g(plim(X))$$

Example:

$$plim(X+Y) = plim(X) + plim(Y)$$

$$plim(\frac{X}{Y}) = \frac{plim(X)}{plim(Y)}\ if\ plim(Y) \neq 0$$

- Base on L.L.N(law of large numbers) and random sample(i.i.d)

$$s_X^2 \xrightarrow{p} = \sigma_X^2 = Var(X)$$

$$s_{xy} \xrightarrow{p} \sigma_{XY} = Cov(X,Y)$$

- then we obtain OLS estimator when $n \longrightarrow \infty$

$$plim\hat{\beta}_1 = plim\left(\frac{s_{xy}}{s_x^2}\right) = \frac{Cov(X_i,Y_i)}{VarX_i}$$

Properties of the OLS estimator: Consistency

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

- then then we could obtain

$$plim\hat{\beta}_1 = \beta_1 \; if \; E[u_i|X_i] = 0$$

- both $\hat{\beta}_0$ and $\hat{\beta}_1$ are **Consistent** on the condition of **Assumption 1**.

Unbiasedness vs Consistency

- *Unbiasedness* & *Consistency* both rely on $E[u_i|X_i] = 0$
- *Unbiasedness* implies that $E[\hat{\beta}_1] = \beta_1$ for a certain sample size n.("small sample")
- *Consistency* implies that the distribution of $\hat{\beta}_1$ becomes more and more *tightly* distributed around β_1 if the sample size n becomes larger and larger.("large sample"")

Sampling Distribution of $\hat{\beta}_0$ and $\hat{\beta}_1$

- Recall: Sampling Distribution of $\overline{Y}$
- Because Y1,...,Yn are i.i.d., then we have

$$E(\overline{Y}) = \mu_Y$$

- Based on the Central Limit theorem(C.L.T), the sample distribution in a large sample can approximates to a normal distribution, thus

$$\overline{Y} \sim N(\mu_Y, \frac{\sigma_Y^2}{n})$$

- the OLS estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ could have similar sample distributions when three least squares assumptions hold.

Sampling Distribution of $\hat{\beta}_0$ and $\hat{\beta}_1$

- Unbiasedness of the OLS estimators implies that

$$E[\hat{\beta}_1] = \beta_1 \; and \; E[\hat{\beta}_0] = \beta_0$$

- Based on the Central Limit theorem(C.L.T), the sample distribution of β in a large sample can approximates to a normal distribution, thus

$$\hat{\beta}_0 \sim N(\beta_0, \sigma^2_{\hat{\beta}_0})$$

$$\hat{\beta}_1 \sim N(\beta_1, \sigma^2_{\hat{\beta}_1})$$

Sampling Distribution of $\hat{\beta}_0$ and $\hat{\beta}_1$ in large-sample

- where it can be shown that

$$\sigma^2_{\hat{\beta}_1} = \frac{1}{n}\frac{Var[(X_i-\mu_x)u_i]}{[Var(X_i)]^2})$$
$$\sigma^2_{\hat{\beta}_0} = \frac{1}{n}\frac{Var(H_iu_i)}{(E[H_i^2])^2})$$

  where

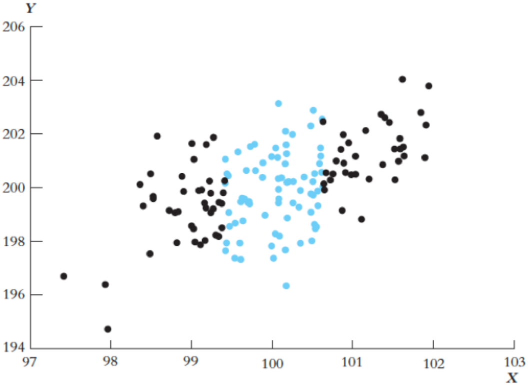
$$H_i = 1 - \left(\frac{\mu_x}{E[X_i^2]}\right)X_i$$

- If $Var(X_i)$ is *small*, it is difficult to obtain an accurate estimate of the effect of X on Y which implies that $Var(\hat{\beta}_1)$ is *large*.

Variation of X

FIGURE 4.6 The Variance of $\hat{\beta}_1$ and the Variance of X

The colored dots represent a set of X_i's with a small variance. The black dots represent a set of X_i's with a large variance. The regression line can be estimated more accurately with the black dots than with the colored dots.



- When more **variation** in X, then there is more information in the data that you can use to fit the regression line.

Summary

Under 3 least squares assumptions, the OLS estimators will be

- **unbiased**
- **consistent**
- **normal sampling distribution**
- *more variation in X, more accurate estimation*

RCT and Simple Regression

- 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.

OLS with Multiple Regressor: Estimation

Violation of the first Least Squares Assumption

- recall simple OLS regression equation

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

- u_i contains all other factors(variables) which potentially affect Y_i.
- Assumption 1 states that they 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?

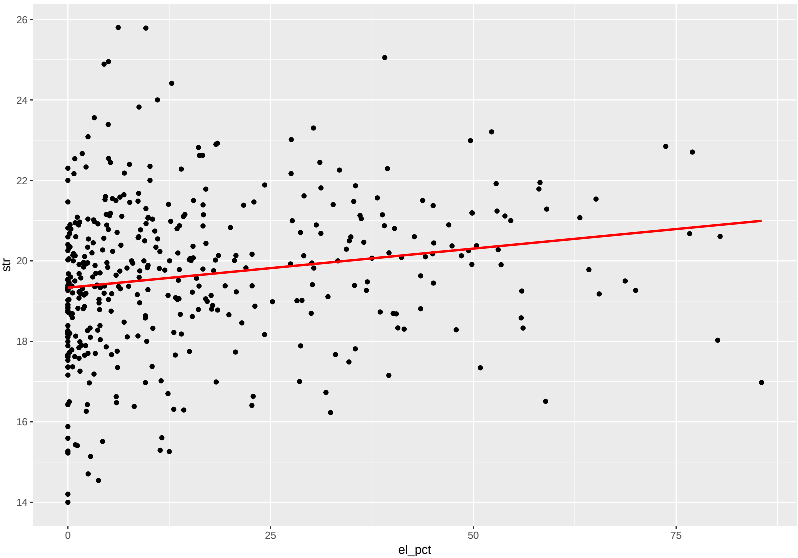
Example: Class Size and Test Score

- one of other factors is the share of immigrants in the class(school,district)
- Suppose that

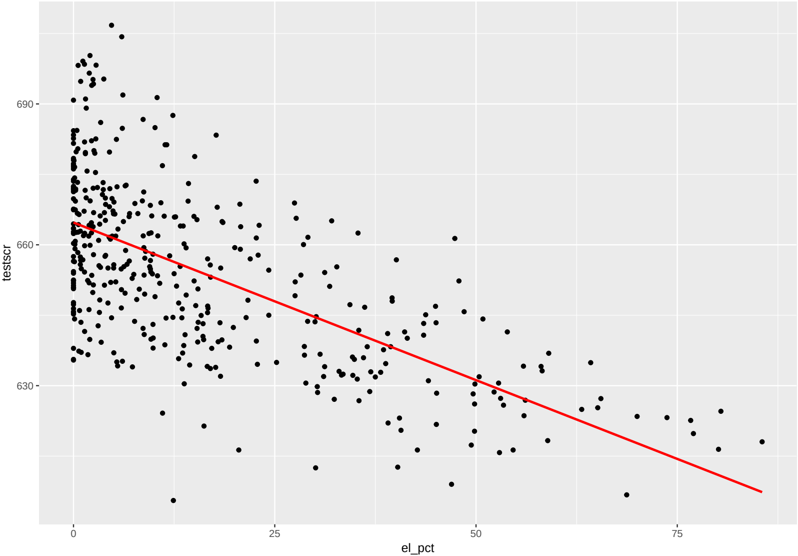
1. small classes have few immigrants(few English learners)
2. large classes have many immigrants(many English learners)

- In this case, class size are correlated with test scores for a fact that class size may be related to percentage of English learners and students who are still learning English likely have lower test scores.
- Which implies that percentage of English learners is contained in u_i, in turn that Assumption 1 is violated.

Scatter plot english learners and STR



Scatter plot english learners and testscr



Omitted Variable Bias(OVB):

- As before, X_i and Y_i represent STR and Test Score.
- Besides, W_i is the share of English learners which we will omit in the regression Thus
- True model:

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

where $E(u_i|X_i, W_i) = 0$ and

- But we can't observe W_i,so we just run the following model

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

where $v_i = \gamma W_i + u_i$

Omitted Variable Bias(OVB): violation of consistency

- we have

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

Omitted Variable Bias(OVB): violation of consistency

- we have

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

- $\hat{\beta}_1$ 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**.

- we have

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

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

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

- If W_i is unrelated to X_i,then $E[\hat{\beta}_1] = \beta_1$
- If W_i is no determinant of Y_i, then it implies also that $E[\hat{\beta}_1] = \beta_1$.
- if both two conditions above are violated *simultaneously*, then $\hat{\beta}_1$ is **biased**.

Omitted Variable Bias(OVB):Directions

- Summary of 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. *Percentage of English learners*
4. *Teachers' Salary*
5. *Family income*

Multiple regression model with k regressors

- 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*
- $\beta_i, j = 1...k$ 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**

Multiple regression model with k regressors

- 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

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

- First order conditions:

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

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

$$\vdots = \vdots$$

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

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

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

Measures of Fit in Multiple Regression

- SER(Standard Error of the Regression) is an estimator of the standard deviation of theu_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 $s_{\hat{u}}^2 = \frac{1}{n-k-1}\sum \hat{u_i^2} = \frac{SSR}{n-k-1}$

- STATA computes the SER but calls it the RMSE.

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 model with k regressors

- 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 ' ' 1
```

Perfect multicollinearity

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

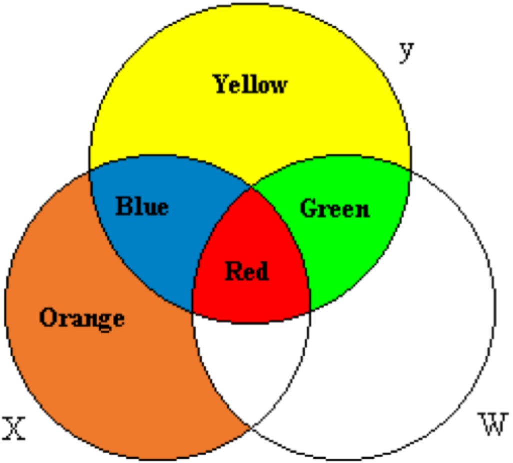
```
##
## Call:
## lm(formula = testscr ~ str + nel_pct + el_pct, data = ca)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -48.845 -10.240  -0.308   9.815  43.461
##
## Coefficients: (1 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 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
## ---
```

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 the least squares assumptions.
- It does not impose theoretical problem for the calculation of OLS estimators.
- If two regressors are highly correlated 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).

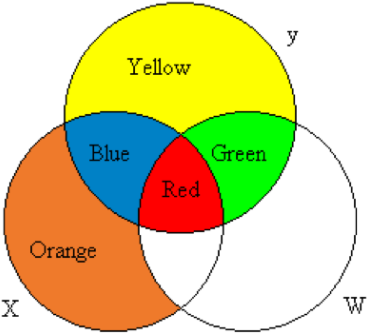


Figure 3a Modest collinearity

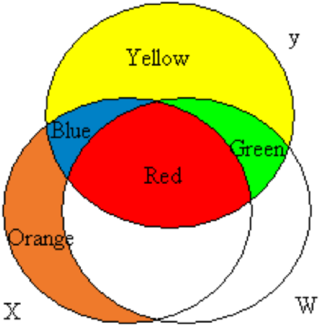


Figure 3b Considerable collinearity

- we use less information (compare the blue and green areas in both figures), the estimation is less precise.

Multiple regression model: class size example

```
##
## ==========================================================================
##                                Dependent variable:
##            --------------------------------------------
##                                      testscr
##                      (1)             (2)             (3)
## --------------------------------------------------------------------------
## str            -2.280 (0.480)  -1.101 (0.380)  -0.069 (0.277)
## el_pct                         -0.650 (0.039)  -0.488 (0.029)
## avginc                                          1.495 (0.075)
## Constant      698.933 (9.467) 686.032 (7.411) 640.315 (5.775)
## --------------------------------------------------------------------------
## Observations        420             420             420
## R2                  0.051           0.426           0.707
## Adjusted R2         0.049           0.424           0.705
## ==========================================================================
## Note:                                                       NA
```

Properties OLS estimators in multiple regression model

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.
- the formal proof need use the knowledge of linear algebra and matrix.We will prove them in a simple case.

Partitioned regression

Partitioned regression: OLS estimator in multiple regression

- 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

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

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 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

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

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

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

- We also have

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

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

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

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

```r
sum(tilde.str)
```

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

```r
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
tilde.strstr <- tilde.str^2
sum(tilde.str_str)
```

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

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

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

Example: Test scores and Student Teacher Ratios(3)

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

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

```r
summary(lm(ca$testscr~tilde.str))
```

```
##
## Call:
## lm(formula = ca$testscr ~ tilde.str)
##
## 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 ***
```

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 $j = 2, 3, ..., k$