ECON 3740: INTRODUCTION TO ECONOMETRICS

INSTRUCTOR: CHAOYI CHEN
Department of Economics and Finance, University of Guelph

Lecture 4

Lecture outline

Last lecture, we learned the types of data, some terminologies for the simple linear regression model, and part of the derivation of OLS estimators. Today, we will

- complete the derivation for the OLS estimators
    - solve for the $\hat{\beta}_0$ & $\hat{\beta}_1$ based on the two FOCs we derived in last lecture

- introduce some measures of fit
    - SST
    - SSE
    - SSR
    - R^2

- change the units measurement
- handle the technique to characterize non-linearity
    - take logs of the dependent variable only
    - take logs of the independent variable only
    - take logs of both

The ordinary least squares estimator (OLS): derivation

- Now, we have two FOCs.

$$1/\text{n}\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = 0$$

$$1/\text{n}\ \sum_{i=1}^{n} x_i\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = 0$$

- We will use these two FOCs to derive β_0 and β_1.
- Step 1: Derive OLS estimator of β_0 as a function of OLS estimator of β_1

$$1/\text{n}\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = 0$$

$$\Longrightarrow 1/n\sum_{i=1}^{n} y_i - 1/n\sum_{i=1}^{n}\hat{\beta}_0 - 1/n\sum_{i=1}^{n}\hat{\beta}_1 x_i = 0$$

$$\Longrightarrow 1/n\sum_{i=1}^{n} y_i - \frac{1}{n}n\hat{\beta}_0 - \hat{\beta}_1\frac{1}{n}\sum_{i=1}^{n} x_i = 0$$

$$\Longrightarrow \bar{y}_i - \hat{\beta}_0 - \hat{\beta}_1\bar{x}_i = 0$$

- The last equation gives

$$\hat{\beta}_0 = \bar{y}_i - \hat{\beta}_1\bar{x}_i$$

The ordinary least squares estimator (OLS): derivation

- Step 2: Solve for the OLS estimator of β_1 ($\hat{\beta}_1$)

$$\textstyle\sum_{i=1}^{n} x_i \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \right) = 0$$

$$\implies \textstyle\sum_{i=1}^{n} x_i \left(y_i - (\bar{y}_i - \hat{\beta}_1 \bar{x}_i) - \hat{\beta}_1 x_i \right) = 0$$

$$\implies \textstyle\sum_{i=1}^{n} x_i \left((y_i - \bar{y}_i) - (\hat{\beta}_1 x_i - \hat{\beta}_1 \bar{x}_i) \right) = 0$$

$$\implies \textstyle\sum_{i=1}^{n} x_i (y_i - \bar{y}_i) - \hat{\beta}_1 \sum_{i=1}^{n} x_i (x_i - \bar{x}_i) = 0$$

Simple Algebra trick can show

$$\implies \textstyle\sum_{i=1}^{n} (x_i - \bar{x}_i)(y_i - \bar{y}_i) - \hat{\beta}_1 \sum_{i=1}^{n} (x_i - \bar{x}_i)(x_i - \bar{x}_i) = 0$$

The ordinary least squares estimator (OLS): derivation

$$\implies \hat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x}_i)(y_i - \bar{y}_i)}{\sum_{i=1}^{n} (x_i - \bar{x}_i)(x_i - \bar{x}_i)} = \frac{S_{xy}}{S_x^2}$$

- Then, $\hat{\beta}_0$ can be calculated by substituting above equation into $\hat{\beta}_0 = \bar{y}_i - \hat{\beta}_1 \bar{x}_i$.
- The OLS predicted values are

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

- The OLS residuals are

$$\hat{\mu}_i = y_i - \hat{y}_i$$

The ordinary least squares estimator (OLS): measures of fit

- How well does the estimated regression line describe the data?
  - Does the regressor X account for much or for little variation in Y ?
  - Are the observations in the scatter plot clustered closely around the regression line?
- Some measures can help us answer above questions!

- **SST**: Total sum of squares measures the total amount of variability in the dependent variable. That is, it measures how spread out the y_i are in the sample.

$$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

- **SSE**: Explained sum of squares measures the sample variation in the $\hat{y}_i$ (where we use the fact that the mean of $\hat{y}_i = \bar{y}$)

$$SSE = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$$

The ordinary least squares estimator (OLS): measures of fit

- **SSR**: Sum of squared residuals measures the total amount of variability that the model does *not* explain

$$SSR = \sum_{i=1}^{n} (\hat{\mu}_i)^2$$

- R^2: measures the variation "explained" by the model

$$R^2 = \frac{ESS}{TSS}$$

- The R^2 ranges from 0 to 1
  - If $R^2 = 0$, x_i explains zero variation in y_i
  - If $R^2 = 1$, x_i explains all of the variation in y_i
  - in practice $R^2 \in (0, 1)$

The ordinary least squares estimator (OLS): The R^2

- Relationship among SST, SSR, SSE?
- The total variation in y can always be expressed as the sum of the explained variation and the unexplained variation SSR. Thus

$$SST = SSE + SSR$$

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n}(\hat{\mu}_i)^2$$

- This implies

$$R^2 = \frac{SSE}{SSR} = \frac{SST - SSR}{SST} = 1 - \frac{SSR}{SST}$$

- At home: prove $SST = SSE + SSR$ mathematically.

The ordinary least squares estimator (OLS): An example for R^2

- Recall the wage problem, our model is

$$wage = \beta_0 + \beta_1 educ + \mu$$

- Once we estimated the model using R, we have following results

```
Call:
lm(formula = wage ~ educ)

Residuals:
    Min      1Q  Median      3Q     Max
-5.3396 -2.1501 -0.9674  1.1921 16.6085

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.90485    0.68497  -1.321    0.187
educ         0.54136    0.05325  10.167   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.378 on 524 degrees of freedom
Multiple R-squared:  0.1648,    Adjusted R-squared:  0.1632
F-statistic: 103.4 on 1 and 524 DF,  p-value: < 2.2e-16
```

The ordinary least squares estimator (OLS): an example for R^2

- Therefore,

$$\hat{wage} = -0.90485 + 0.54136\ educ\ \ \&\ R^2 = 0.1648$$

- The education level explains 16.5% of the variation in wage based on our sample.
- Even the R^2 in above example is high, generally speaking, in the social sciences, we have to notice that low R^2 in regression equations are not uncommon, especially for cross-sectional analysis. We will discuss this issue more generally under multiple regression analysis. However, it is worth emphasizing now that a seemingly low R^2 does **not** necessarily mean that an OLS regression equation is useless.

The ordinary least squares estimator (OLS): changing units of measurement

- Data Scaling
  - Predictions in different units
  - Different interpretations
- Example:

$$\text{wage}{=}\beta_0 + \beta_1 educ + \mu$$
$$educ \text{ is in years}$$
$$wage \text{ is in dollar}$$

- Estimates:

$$\hat{wage} = \hat{\beta}_0 + \hat{\beta}_1 educ$$

- Intuitively, the result implies one more year education increases $\hat{\beta}_1$ dollars in wage.

The ordinary least squares estimator (OLS): changing units of measurement

- Now, suppose the wage unit changes from dollars to cents,

$$\implies \quad \hat{wage}_{dollars} = \frac{1}{100}\hat{wage}_{cents} \tag{1}$$

- Recall, the original estimation gives

$$\hat{wage} = \hat{\beta}_0 + \hat{\beta}_1 educ \tag{2}$$

- next, we substitute equation (1) in to (2)

$$\frac{1}{100}\hat{wage}_{cents} = \hat{\beta}_0 + \hat{\beta}_1 educ \tag{3}$$

$$\hat{wage}_{cents} = 100\hat{\beta}_0 + 100\hat{\beta}_1 educ$$

- Intuitively, the result implies one more year education increases $100\hat{\beta}_1$ cents in wage.

The ordinary least squares estimator (OLS): handling non-linearity

- In the real life, you will often encounter regression equations where the dependent variable appears in logarithmic form. Why we need it? Recall the wage-education case, where we regressed dollars of wage on years of education. Because of the linear nature, $\hat{\beta}_1$ dollars is the increase for either the first year of education or the tenth year of education. This can be highly doubted!
- Ideally, a better characterization of how wage changes with education is that each year of education increases wage by a constant *percentage*.
- Three ways to handle non-linearity
  - Take logs of the dependent variable only
  - Take logs of the independent variable only
  - Take logs of both

The ordinary least squares estimator (OLS): handling non-linearity

- Case 1: *wage* is in logs (take logs of the dependent variable only)
$$log(\hat{wage}) = \hat{\beta}_0 + \hat{\beta}_1 educ$$

- How do we interpret $\hat{\beta}_1$?
- Totally differentiate gives
$$\frac{1}{\hat{wage}} \partial \hat{wage} = \hat{\beta}_1 \partial educ$$

- To simplfy, we have
$$\underbrace{\frac{\partial \hat{wage}}{\hat{wage}} \times 100}_{\%\ change} = (\hat{\beta}_1 \times 100) \underbrace{\partial educ}_{unit\ \ change}$$

- Therefore, we can interpret the story as a one-year increase in education tields a $\hat{\beta}_1$% increase in wage