ECON 3740: INTRODUCTION TO ECONOMETRICS

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

Lecture 6

Lecture outline

Last lecture, we learned, under four assumptions, the OLS estimators are unbiased. We also studied the difference between homoskedastic error and heteroskedastic error. Today, we will

- derive the variance of $\hat{\beta}_1$
- estimate the error variance
- study regression through the origin and regression on a constant
- summary for topic four: The simple Regression Model

Simple linear regression model: sample variance of the OLS estimators

- At the end of the last lecture, we showed that, under assumptions 1-5, the variance of $\hat{\beta}_1$ is

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

- Now, let's go through the derivation formally.
- Recall, from last lecture,

$$\begin{aligned}\hat{\beta}_1 &= \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}_i\right)\left(\beta_1(X_i - \bar{X}) + (\mu_i - \bar{\mu})\right)}{\sum_{i=1}^{n}\left(x_i - \bar{x}_i\right)\left(x_i - \bar{x}_i\right)} \\ &= \beta_1 + \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}_i\right)\mu_i}{\sum_{i=1}^{n}\left(x_i - \bar{x}_i\right)\left(x_i - \bar{x}_i\right)}\end{aligned}$$

Simple linear regression model: sample variance of the OLS estimators

- Hence,

$$Var(\hat{\beta}_1) = Var(\beta_1 + \frac{\sum_{i=1}^{n}(x_i - \bar{x}_i)\mu_i}{\sum_{i=1}^{n}(x_i - \bar{x}_i)(x_i - \bar{x}_i)}) = Var(\frac{\sum_{i=1}^{n}(x_i - \bar{x}_i)\mu_i}{\sum_{i=1}^{n}(x_i - \bar{x}_i)(x_i - \bar{x}_i)})$$

$$= (\frac{1}{\sum_{i=1}^{n}(x_i - \bar{x}_i)^2})^2 Var(\sum_{i=1}^{n}(x_i - \bar{x}_i)\mu_i)$$

- Note that the μ_i are independent random variables across i, which implies the variance of the sum is the sum of the variances. Hence,

$$= (\frac{1}{\sum_{i=1}^{n}(x_i - \bar{x}_i)^2})^2 \Big(\sum_{i=1}^{n}(x_i - \bar{x}_i)^2 Var(\mu_i)\Big)$$

Simple linear regression model: sample variance of the OLS estimators

- If the error is *heteroskedastic*, we cannot simplify further due to the fact that μ_i changes as x_i changes.
- However, with assumption 5 (the homoskedastic error assumption), $Var(\mu_i) = \sigma^2$, which is a constant for all i. Therefore, we can simplify the variance as

$$Var(\hat{\beta}_1) = \left(\frac{1}{\sum_{i=1}^{n}\left(x_i - \bar{x}_i\right)^2}\right)^2 \sigma^2 \left(\sum_{i=1}^{n}\left(x_i - \bar{x}_i\right)^2\right)$$

$$= \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

- From the above expression, we can see the variance depends on the error variance, σ^2, and the totoal variation in x_i, the SST_x.

Simple linear regression model: sample variance of the OLS estimators

- Relationships and intuitions:
  - 1. The larger the error variance, the larger is $Var(\hat{\beta}_1)$. Intuitively, more variation in the unobservables affecting y makes it more difficult to precisely estimate β_1.
  - 2. As the variability in the x_i increases, the variance of $\hat{\beta}_1$ decreases. Intuitively, the more spread out is the sample of independent variables, the easier it is to trace out the relationship between $E(y|x)$ and x. Therefore, the easier it is to estimate β_1. 1 decreases.

Simple linear regression model: estimate error variance

- With the data, we can compute SST_x.
- However, we do **not** know σ^2 in general.
- To get the variance $Var(\hat{\beta}_1)$, we need to estimate the error variance first.

$$y_i = \beta_0 + \beta_1 x_i + \mu_i$$
$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{\mu}_i$$

- The first equation is the the population model in terms of a randomly sampled observation
- The second express y_i in terms of its fitted value and residual
- Comparing two equations, we see that the error shows up in the equation containing the *population* parameters
- On the other hand, the residuals show up in the *estimated* equation

Simple linear regression model: estimate error variance

- If we know the error μ_i, then we can use $\frac{1}{n}\sum_{i=1}^{n}\mu_i^2$ as an unbiased estimator for $Var(\mu) = \sigma^2$.
- Unfortunately, this is not a true estimator, because we do not observe the errors μ_i. But, the good news is we do have estimates of the μ_i, the OLS residuals $\hat{\mu}_i$.
- However, if we simply replace μ_i with $\hat{\mu}_i$ in $\frac{1}{n}\sum_{i=1}^{n}\mu_i^2$, the estimator $(\frac{1}{n}\sum_{i=1}^{n}\hat{\mu}_i^2)$ turns out to be a *biased* estimator. (Prove it at home).
- Intuitively, the biasedness is essentially because it does not account for two restrictions that must be satisfied by the OLS residuals.

$$\sum_{i}^{n}\hat{\mu}_i = 0 \qquad \sum_{i}^{n}x_i\hat{\mu}_i = 0$$

- Why the two restrictions matter? if we know $n-2$ of the residuals. Then, based on the above two equation, we can always get the other two residuals (2 unknown, 2 equations).Thus, there are only $n-2$ degrees of freedom in $\hat{\mu}_i$ istead of n degrees of freedom in μ_i.

Simple linear regression model: estimate error variance

- Therefore, an unbiased estimator of σ^2 that we will use makes a degrees of freedom adjustment is

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

- Then, by replacing σ^2 by $\hat{\sigma}^2$, we can show that $\frac{\hat{\sigma}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$ is an unbiased estimator of $Var(\hat{\beta}_1)$
- $se(\hat{\beta}_1) = \frac{\hat{\sigma}}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$ is the estimator of the standard deviations of $\hat{\beta}_1$, namely, standard error of $\hat{\beta}_1$.
- Similarly, we can obtain the unbiased estimator $Var(\hat{\beta}_0)$ and the standard error of $\hat{\beta}_0$

Simple linear regression model: study regression through the origin and regression on a constant

- Now, let's talk about a special case of the simple linear regression model by imposing the restriction that, when $x = 0$, the expected value of y is zero. Hence, The population model is $y_i = \beta_1 x_i + \mu_i$.
- Note that the population line has to be through the origin point $(0, 0)$.
- Why this is reasonable and useful? Example: if income (x) is zero, then income tax revenues (y) must also be zero.
- To estimate the slope (we do not have intercept in this case), the OLS method requires us to solve the following minimization plroblem

$$\underset{\hat{\beta}_1}{Min} \sum_{i=1}^{n} (y_i - \hat{\beta}_1 x_i)^2$$

Simple linear regression model: study regression through the origin and regression on a constant

- Differentiate the objective function w.r.t $\hat{\beta}_1$, we have the FOC

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

- Thus, the OLS estimator of β_1

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$$

- The residuals are $\hat{\mu}_i = y_i - \hat{\beta}_1 x_i$ and the $R^2 = 1 - \frac{\sum_{i=1}^{n} \hat{\mu}_i}{SST}$
- Note that, in this typical case, $SST = \sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2$

Simple linear regression model: study regression through the origin and regression on a constant

- Another interesting problem is what happens if we only regress on a constant?
- That is, we set the slope to zero (which means we need not even have an x) and estimate an intercept only?

- The answer is simple: the estimator of the intercept (β_0) is $\bar{y}$.
- This fact is usually shown in basic statistics, where it is shown that the constant that produces the smallest sum of squared deviations is always the sample average.

Simple linear regression model: summary

- In this topic, we have introduced the simple linear regression model.
- We have learned using OLS method to derive the estimators of the slope and the intercept parameters.
- We have demonstrated the algebra of the OLS regression line, including computation of fitted values and residuals, and the obtaining of predicted changes in the dependent variable for a given change in the independent variable.
- We have discussed two issues of practical importance: (1) change the units of measurement (2) the use of the natural log to model nonlinearity.

Simple linear regression model: summary

- We showed that under four assumptions, the OLS estimators are unbiased. (The expected value of the OLS estimator equals to the true parameter)

- By imposing one more assumption, "homoskedastic error", we showed the simpler expressions of the variance of the OLS estimators. As we saw, the variance of the slope estimator $\hat{\beta}_1$ increases as the error variance increases, and it decreases when there is more sample variation in the independent variable.

- We showed an unbiased estimator of the σ^2, $\hat{\sigma}^2 = \frac{SSE}{(n-2)}$.Then, we use $\hat{\sigma}^2$ to get the unbiased estimator of $Var(\hat{\beta}_1)$.

- At last, we briefly discussed regression through the origin and regression on a constant