Bivariate OLS Regression (Part 2 of 2)

Lecture 4

Recap: Properties of OLS

- Under special circumstances, the OLS slope estimate will be a "good" estimate of the causal relationship between X and Y.
- Need two things:
  1. The true parameters must be causal
  2. OLS must provide a "good" estimate of the true values
- To understand what "good" means, first remember that $\hat{\beta}_0$ and $\hat{\beta}_1$ are themselves random variables, not constants
- A "good" estimate is unbiased and precise

Bivariate OLS: Precision

Bivariate OLS Assumptions

1. Linear in parameters
2. Random sampling: (x_i, y_i) are i.i.d.
3. Sample variation in the explanatory variable
4. Zero conditional mean: $E(u_i|x_i) = 0$
5. Homoskedasticity: $Var(u_i|x_i) = \sigma_u^2$

- We learned (4) is important
- We will focus on (5) today

Gauss-Markov Theorem

- If the error term is homoskedastic, then OLS is even better
  - Note: Homoskedasticity is not required for OLS to be unbiased
- Assumptions 1-5 combined are sufficient for OLS to be BLUE
  - Best (most efficient, most accurate)
  - Linear (as opposed to nonlinear, e.g., maximum likelihood)
  - conditionally Unbiased
  - Estimator
- In sum, of all linear unbiased estimators, OLS provides the most precision(= "good")

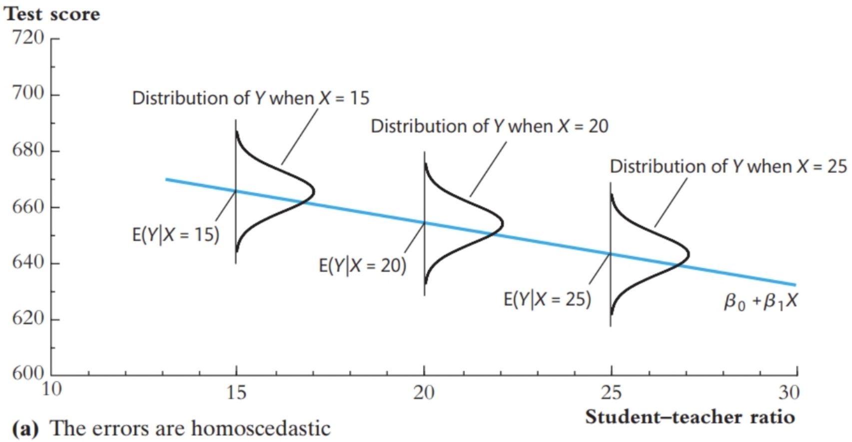
Homoskedasticity

- It is possible for $E(u|X) = 0$ but for $Var(u|X)$ to vary with X
- Homoskedasticity: The error u has the same variance given any value of the explanatory variable

$$Var(u_i|x_i) = \sigma_u^2$$

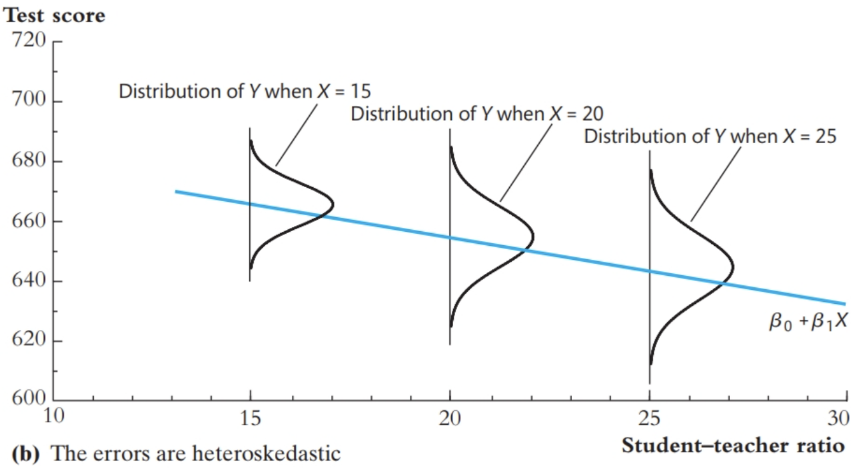
  - The square root of σ_u^2, σ_u, is the standard deviation of the error
  - A larger σ_u means that the distribution of the unobservables affecting y is more spread out
- If the variance of u varies at different levels of X, we call the error term heteroskedastic

Example of Homoskedasticity



Source: Stock and Watson Ch.5 Figure5-2

Example of Heteroskedasticity

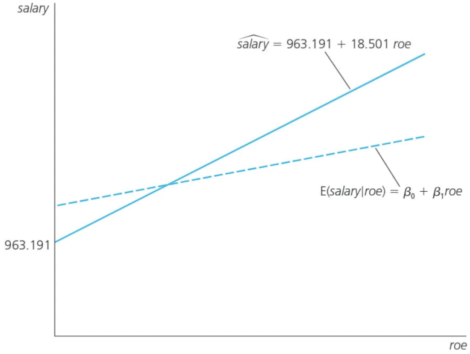


Source: Stock and Watson Ch.5 Figure5-2

Let's get Precise about Precision

Example: CEO salary and return on equity

The OLS regression line and the (unknown) population regression function



Fitted regression line depends on a sample, so if we use another sample from a different data, it will give a different regression line.

Source: Wooldridge Ch.2 Figure2-5

Variation in OLS estimates

- How far can we expect $\hat{\beta}_1$ to be away from β_1 on average? (=sampling variability)
  - Sampling variability is measured by the estimator's variances

Recap: Sampling

- Statistical parameters are estimates of a true "population" parameter
- We learn about the true population parameter by sampling individuals from the population
- Depending on the sample, the estimates will be nearer or farther away from the true population values
  - Notation:
    - β: population parameter
    - $\hat{\beta}$: estimate based on finite data
    - $Var(\hat{\beta}_1)$ (or $\sigma^2_{\hat{\beta}_1}$): variance of estimate

Variation in OLS estimates

- When n large, distribution of $\hat{\beta}_1$ is approximated by the normal distribution:

$$\hat{\beta}_1 \sim N(\beta_1, Var(\hat{\beta}_1))$$

  - Application of the central limit theorem
  - Enables analytical hypothesis testing

Variation in OLS estimates

- Under OLS assumptions Additional steps shown in bonus slides :

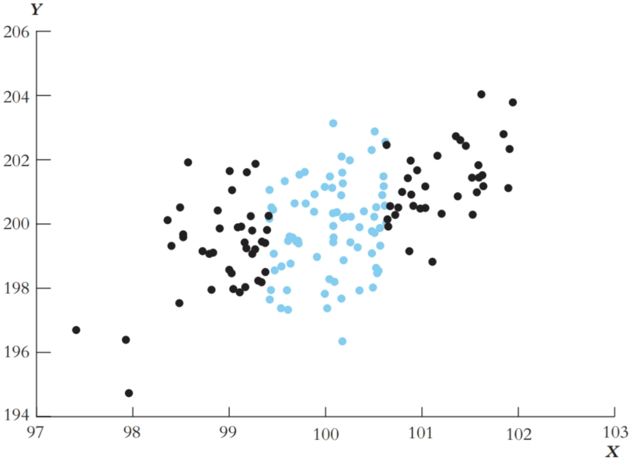
$$Var(\hat{\beta}_1) = \frac{\sigma_u^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sigma_u^2}{\text{TSS}_x} = \frac{\sigma_u^2}{N\sigma_x^2}$$

  - $\text{TSS}_x = \sum_{i=1}^{n}(x_i - \bar{x})^2$ is the *total sum of squares* (=total variation in x)
  - $\sigma_{\hat{\beta}_1} = \frac{1}{\sqrt{n}}\frac{\sigma_u}{\sigma_x}$

- The precision of the OLS estimate is greater as:
  1. σ_u^2 falls
  2. the variance of x_i rises
     - Intuition: you need to observe y for a greater range of x to get more information
  3. the sample size n rises

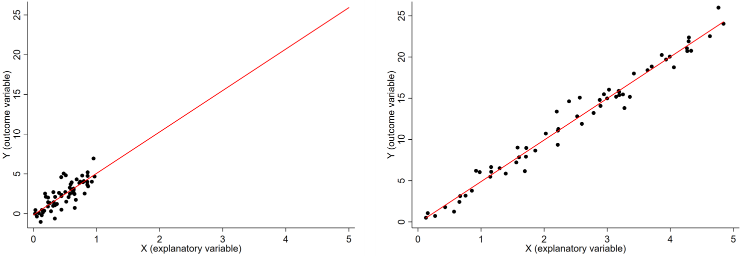
Precision of $\hat{\beta}_1$ and variance of X

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.



(Figure 4.5 in Stock and Watson)

Precision of $\hat{\beta}_1$ and variance of X



- Same true β, but more variance in X on right
- Which one is more precise?

Estimating the error variance

- Remember, the errors are never observed, while the residuals are computed from the data (the OLS residuals $\hat{u_i}$)
- We can use data to estimate σ_u^2

$$\hat{\sigma_u^2} = \frac{1}{n-2}\sum_{i=1}^{n} \hat{u_i}^2$$

$n-2$ is a degrees of freedom adjustment(n is the number of observations, 2 is the number of estimated regression coefficients)

- Under OLS assumptions (1)-(5), $E(\hat{\sigma_u^2}) = \sigma_u^2$

Standard Error for $\hat{\beta}_1$

- Calculation of standard error for $\hat{\beta}_1$:

$$SE(\hat{\beta}_1) \text{ (or } \hat{\sigma}_{\hat{\beta}_1}) = \sqrt{\hat{Var}(\hat{\beta}_1)} = \sqrt{\hat{\sigma_u^2}/\text{TSS}_x} = \hat{\sigma_u}/\sqrt{\text{TSS}_x}$$

- The estimated standard deviations of the regression coefficients are called "standard errors."
  - They measure how precisely the regression coefficients are estimated

Bivariate OLS: Inference

Main inference tools

Generally, we want to do statistical inference about hypotheses relating to the population parameters β.

- H_0: $\beta_1 = \theta$ vs. $H_A : \beta_1 \neq \theta$ (two-sided)
- H_0: $\beta_1 = \theta$ vs. $H_A : \beta_1 < \theta$ (or $\beta_1 > \theta$) (one-sided)

Note: θ is hypothesized value of the coefficient

Generally, two-sided tests are more appropriate (use as default)

t-statistics

- Recall that if W is a normal variable, then subtracting its mean and dividing by its standard deviation yields a standard normal variable

$$W \sim N(\mu_W, \sigma_W^2) \rightarrow \frac{W - \mu_W}{\sigma_W} \sim N(0,1)$$

- If the standard deviation is estimated instead of known, then normalizing yields a t-distribution

$$W \sim N(\mu_W, \sigma_W^2) \rightarrow \frac{W - \mu_W}{\hat{\sigma}_W} \sim t_{n-k-1}$$

- The t-distribution is close to the standard normal distribution if $n - k - 1$ is large
- In a bivariate regression, the degrees of freedom is n-2, since there are two parameters estimated

t-statistics

- So applying this to $\hat{\beta}_1$:

$$\hat{\beta}_1 \sim N(\beta_1, \sigma^2_{\hat{\beta}_1}) \rightarrow \frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}_{\hat{\beta}_1}} \sim t_{n-k-1}$$

- If the t-stat is "small", can't reject null

"Statistically significant"

- Could the true $\hat{\beta}$ be zero, i.e. no effect of X on Y?
- If a regression coefficient is different from zero in a two-sided test, the corresponding variable is said to be "statistically significant"
- If the number of degrees of freedom is large enough so that the normal approximation applies, the following rules of thumb apply:

$|t| > 1.645 \rightarrow$ statistically significant at 10% level

$|t| > 1.96 \rightarrow$ statistically significant at 5% level

$|t| > 2.576 \rightarrow$ statistically significant at 1% level

t-statistics

- Example: $\hat{\beta}_1 = 0.084$ and $SE(\hat{\beta}_1) = 0.014$
- The default test in Stata/R is against a zero ("no effect") null: t-test of H_0: $\beta_1 = 0$

  $\frac{0.084-0}{0.014} = 6.0$
- Note: we can technically do the same test against any other null hypothesis, i.e. t-test of H_0: $\beta_1 = 0.1$

  $\frac{0.084-0.1}{0.014} = -1.14$

p-values

- p-value is the probability of obtaining a $\hat{\beta}$ at least as far away from the null as the observed outcome, if the null is true

$$\text{p-value} \approx 2\Phi(-|t_{n-k-1}|)$$

- The null hypothesis is rejected if the p-value is smaller than the significance level
- In the example with $df = 40$ and $t = 1.85$, the p-value is computed as
  - p-value $\approx 2\Phi(-|1.85|) = 0.0718$
  - R code: `2*pt(-abs(1.85),df=40)`

Example: CEO salary and return on equity

```
Call:
lm(formula = salary ~ roe, data = data)

Residuals:
    Min      1Q  Median      3Q     Max
-1160.2  -526.0  -254.0   138.8 13499.9

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   963.19     213.24   4.517 1.05e-05 ***
roe            18.50      11.12   1.663   0.0978 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1367 on 207 degrees of freedom
Multiple R-squared:  0.01319,	Adjusted R-squared:  0.008421
F-statistic: 2.767 on 1 and 207 DF,  p-value: 0.09777
```

- R automatically reports $H_0 : \beta_1 = 0$

$$\frac{18.50 - 0}{11.12} = 1.663$$

- This tests whether the coefficient estimate is "statistically significantly different from zero"
- R code for p-values: 2*pt(-abs(1.663),df=207)=0.0978

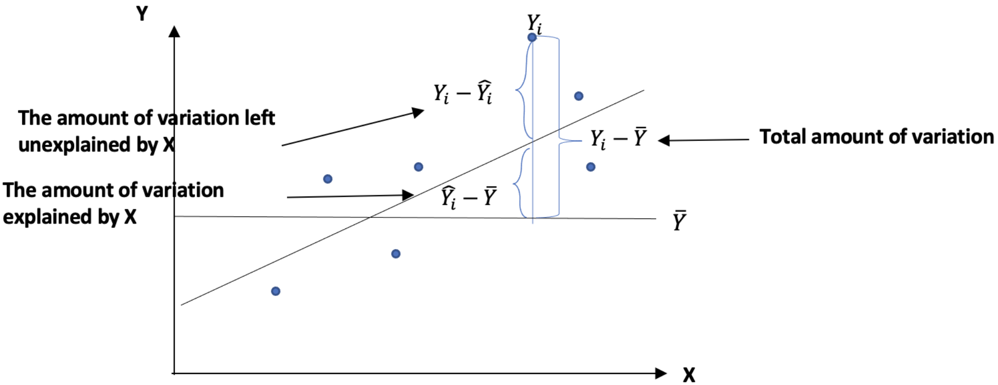
Confidence intervals

- $\hat{\beta}_1 = 0.084$ and $SE(\hat{\beta}_1) = 0.014$
- 95% confidence interval for β_1 is $[\hat{\beta}_1 - t_{\alpha/2,n-k-1} * SE(\hat{\beta}_1), \hat{\beta}_1 + t_{\alpha/2,n-k-1} * SE(\hat{\beta}_1)] = [0.0567, 0.1114]$
- For $(1-\alpha)$% of possible samples, the confidence interval (CI) will include the true value
- For any θ inside of this CI, you would not reject H_0: $\beta_1 = \theta$ at the α% level

Bivariate OLS: Goodness of Fit

Measure of Fit: R-squared

- $\mathrm{TSS}_Y = \sum_{i=1}^{n} \left(Y_i - \bar{Y}\right)^2$ is the *total sum of squares*
  - TSS_Y is the total amount of variation in the sample
- $\mathrm{ESS} = \sum_{i=1}^{n} \left(\hat{Y}_i - \bar{Y}\right)^2$ is the *explained sum of squares*
  - ESS is the amount of variation explained by X
- $\mathrm{SSR} = \sum_{i=1}^{n} \hat{u}_i^2$ is the *sum of squared residuals*.
  - SSR is the amount of variation left unexplained by X
- Note that $\mathrm{TSS} = \mathrm{ESS} + \mathrm{SSR}$


Y
Y_i
$Y_i - \hat{Y}_i$
The amount of variation left unexplained by X
$Y_i - \bar{Y}$
Total amount of variation
The amount of variation explained by X
$\hat{Y}_i - \bar{Y}$
$\bar{Y}$
X

Measure of Fit: R-squared

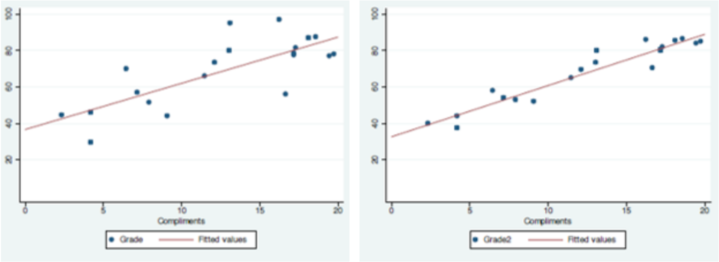
The R^2 is the fraction of the sample variation in Y_i that is explained by X_i

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

The R^2 ranges between 0 and 1:

- $R^2 = 1$ if and only if $\mathrm{SSR} = 0$
- $R^2 = 0$ if and only if $\mathrm{ESS} = 0$
- A higher value indicates a better "goodness of fit" or a larger share of the variation explained

Goodness of Fit Graph



- Right is same data as left with shrunken error (decreased var of u)
- Right will have higher R^2 – larger fraction of variation in Y is explained by X

Example: CEO salary and return on equity

```
Call:
lm(formula = salary ~ roe, data = data)

Residuals:
    Min      1Q  Median      3Q     Max 
-1160.2  -526.0  -254.0   138.8 13499.9 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   963.19     213.24   4.517 1.05e-05 ***
roe            18.50      11.12   1.663   0.0978 .  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1367 on 207 degrees of freedom
Multiple R-squared:  0.01319,	Adjusted R-squared:  0.008421 
F-statistic: 2.767 on 1 and 207 DF,  p-value: 0.09777
```

- How much of the variation in salary is explained by the return on equity? Interpret R^2.

Practice Questions

```
> summary(lm(earnings~training,data=data))

Call:
lm(formula = earnings ~ training, data = data)

Residuals:
   Min     1Q Median     3Q    Max 
-6.349 -4.555 -1.829  2.917 53.959 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   4.5548     0.4080  11.162  < 2e-16 ***
training      1.7943     0.6329   2.835  0.00479 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.58 on 443 degrees of freedom
Multiple R-squared:  0.01782,	Adjusted R-squared:  0.01561 
F-statistic: 8.039 on 1 and 443 DF,  p-value: 0.004788
```

We are interested in effects of a job training program, where men with poor labor market histories were randomly assigned to control and treatment group. $earnings$ is earnings in thousands of dollars; $training$ is the training assignment indicator

1. What are average earnings for men with no training? What is the average effect of training on earnings?
2. Construct a 95% confidence interval for $\hat{\beta}_1$
3. How much of the variation in earnings is explained by training?
4. Do you think the zero conditional mean assumption is satisfied? Why?

Bonus slides

$Var(\hat{\beta}_1)$

First, using (1) "linear in parameter assumption", we can rewrite $\hat{\beta}$ as follows:

$$\hat{\beta}_1 = \frac{\sum(x_i - \overline{x})(y_i - \overline{y})}{\sum(x_i - \overline{x})^2} = \frac{\sum(x_i - \overline{x})y_i}{\sum(x_i - \overline{x})^2}$$

$$= \frac{\sum(x_i - \overline{x})(\beta_0 + \beta_1 x_i + u_i)}{\sum(x_i - \overline{x})^2}$$

$$= \frac{\beta_0 \sum(x_i - \overline{x}) + \beta_1 \sum(x_i - \overline{x})x_i + \sum(x_i - \overline{x})u_i}{\sum(x_i - \overline{x})^2}$$

$$= \frac{0 + \beta_1 \sum(x_i - \overline{x})^2 + \sum(x_i - \overline{x})u_i}{\sum(x_i - \overline{x})^2} = \beta_1 + \frac{\sum(x_i - \overline{x})u_i}{\sum(x_i - \overline{x})^2}$$

Algebra Notes:

Line 1: $\sum(x_i - \overline{x})(y_i - \overline{y}) = \sum(x_i - \overline{x})y_i + \overline{y}\sum(x_i - \overline{x}) = \sum(x_i - \overline{x})y_i + \overline{y}(N\overline{x} - N\overline{x}) = \sum(x_i - \overline{x})y_i$

Line 4: $\sum(x_i - \overline{x})x_i = \sum(x_i^2 - \overline{x}x_i) = \sum(x_i^2) - N\overline{x}^2 = \sum(x_i - \overline{x})^2$ (see Lecture 3 slide 46 footnote)

$Var(\hat{\beta}_1)$

Now, taking the variance, and using (2) random sampling and (5) homoskedasticity assumptions:

$$
\begin{aligned}
Var(\hat{\beta}) &= Var\left(\beta_1 + \frac{\sum(x_i - \overline{x})u_i}{\sum(x_i - \overline{x})^2}\right) = Var\left(\frac{\sum(x_i - \overline{x})u_i}{\sum(x_i - \overline{x})^2}\right) \\
&= \frac{1}{(\sum(x_i - \overline{x})^2)^2} Var\left(\sum(x_i - \overline{x})u_i\right) = \frac{1}{(\sum(x_i - \overline{x})^2)^2} \sum Var\left((x_i - \overline{x})u_i\right) \\
&= \frac{1}{(\sum(x_i - \overline{x})^2)^2} \sum(x_i - \overline{x})^2 Var(u_i) = \frac{1}{\sum(x_i - \overline{x})^2} \sigma_u^2
\end{aligned}
$$

Note: We "condition" on the x_i, i.e. we have our specific sample, so we can treat x_i as non-random.

Back