Econ 422 – Lecture Notes
Part II

(These notes are slightly modified versions of lecture notes provided by Stock and Watson, 2007. They are for instructional purposes only and are not to be distributed outside of the classroom.)

Regression with a Single Regressor: Hypothesis Tests and Confidence Intervals

Overview

- Now that we have the sampling distribution of OLS estimator, we are ready to perform hypothesis tests about β_1 and to construct confidence intervals about β_1
- Also, we will cover some loose ends about regression:
  - Regression when X is binary (0/1)
  - Heteroskedasticity and homoskedasticity
  - Efficiency of the OLS estimator
  - Use of the t-statistic in hypothesis testing

But first… a big picture view (and review)

We want to learn about the slope of the population regression line, using data from a sample (so there is sampling uncertainty). There are four steps towards this goal:

1. State precisely the population object of interest
2. Derive the sampling distribution of an estimator (this requires certain assumptions)
3. Estimate the variance of the sampling distribution (which the CLT tells us is all you need to know if *n* is large) – that is, finding the standard error (*SE*) of the estimator – *using only the information in the sample at hand!*
4. Use the estimator ($\hat{\beta}_1$) to obtain a point estimate and, with its *SE*, hypothesis tests, and confidence intervals.

Object of interest: β_1 in,

$$Y_i = \beta_0 + \beta_1 X_i + u_i,\ i = 1,\ldots,\ n$$

$\beta_1 = \Delta Y/\Delta X$, for an autonomous change in X (***causal effect***)

The Least Squares Assumptions:

1. $E(u|X = x) = 0$.
2. (X_i,Y_i), $i = 1,\ldots,n$, are i.i.d.
3. Large outliers are rare ($E(X^4) < \infty$, $E(Y^4) < \infty$.

The Sampling Distribution of $\hat{\beta}_1$:

Under the LSA's, for n large, $\hat{\beta}_1$ is approximately distributed,

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma_v^2}{n\sigma_X^4}\right), \text{ where } v_i = (X_i - \mu_X)u_i$$

Hypothesis Testing and the Standard Error of $\hat{\beta}_1$

The objective is to test a hypothesis, like $\beta_1 = 0$, using data – to reach a tentative conclusion whether the (null) hypothesis is correct or incorrect.

General setup

Null hypothesis and **two-sided** alternative:

$$H_0: \ \beta_1 = \beta_{1,0} \ \text{ vs. } H_1: \ \beta_1 \neq \beta_{1,0}$$

where $\beta_{1,0}$ is the hypothesized value under the null.

Null hypothesis and **one-sided** alternative:

$$H_0: \ \beta_1 = \beta_{1,0} \ \text{ vs. } H_1: \ \beta_1 < \beta_{1,0}$$

General approach: construct *t*-statistic, and compute *p*-value (or compare to $N(0,1)$ critical value)

- In general: $$t = \frac{\text{estimator - hypothesized value}}{\text{standard error of the estimator}}$$

  where the *SE* of the estimator is the square root of an estimator of the variance of the estimator.

- ***For testing the mean of Y***: $$t = \frac{\bar{Y} - \mu_{Y,0}}{s_Y / \sqrt{n}}$$

- ***For testing β_1***, $$t = \frac{\hat{\beta}_1 - \beta_{1,0}}{SE(\hat{\beta}_1)},$$

  where $SE(\hat{\beta}_1)$ = the square root of an estimator of the variance of the sampling distribution of $\hat{\beta}_1$

Formula for SE($\hat{\beta}_1$)

Recall the expression for the variance of $\hat{\beta}_1$ (large n):

$$\text{var}(\hat{\beta}_1) = \frac{\text{var}[(X_i - \mu_x)u_i]}{n(\sigma_X^2)^2} = \frac{\sigma_v^2}{n\sigma_X^4}, \text{ where } v_i = (X_i - \mu_X)u_i.$$

The estimator of the variance of $\hat{\beta}_1$ replaces the unknown population values of σ_v^2 and σ_X^4 by estimators constructed from the data:

$$\hat{\sigma}^2_{\hat{\beta}_1} = \frac{1}{n} \times \frac{\text{estimator of } \sigma_v^2}{(\text{estimator of } \sigma_X^2)^2} = \frac{1}{n} \times \frac{\frac{1}{n-2}\sum_{i=1}^{n} \hat{v}_i^2}{\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right]^2}$$

where $\hat{v}_i = (X_i - \bar{X})\hat{u}_i$.

$$\hat{\sigma}^2_{\hat{\beta}_1} = \frac{1}{n} \times \frac{\frac{1}{n-2}\sum_{i=1}^{n} \hat{v}_i^2}{\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right]^2}, \text{ where } \hat{v}_i = (X_i - \bar{X})\hat{u}_i .$$

$$SE(\hat{\beta}_1) = \sqrt{\hat{\sigma}^2_{\hat{\beta}_1}} = \text{the standard error of } \hat{\beta}_1$$

OK, this is a bit nasty, but:

- It is less complicated than it seems. The numerator estimates var(v), the denominator estimates var(X).
- Why the degrees-of-freedom adjustment $n - 2$? Because two coefficients have been estimated (β_0 and β_1).
- $SE(\hat{\beta}_1)$ is computed by regression software
- STATA has memorized this formula so you don't need to.

Summary: To test H_0: $\beta_1 = \beta_{1,0}$ v. H_1: $\beta_1 \neq \beta_{1,0}$,

- Construct the *t*-statistic

$$t = \frac{\hat{\beta}_1 - \beta_{1,0}}{SE(\hat{\beta}_1)} = \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\hat{\sigma}^2_{\hat{\beta}_1}}}$$

- Reject at 5% significance level if |t| > 1.96
- The *p*-value is $p = \Pr[|t| > |t^{act}|]$ = probability in tails of normal outside $|t^{act}|$; you reject at the 5% significance level if the *p*-value is < 5%.
- This procedure relies on the large-*n* approximation; typically $n = 50$ is large enough for the approximation to be excellent.

Example: *Test Scores* and *STR*, California data

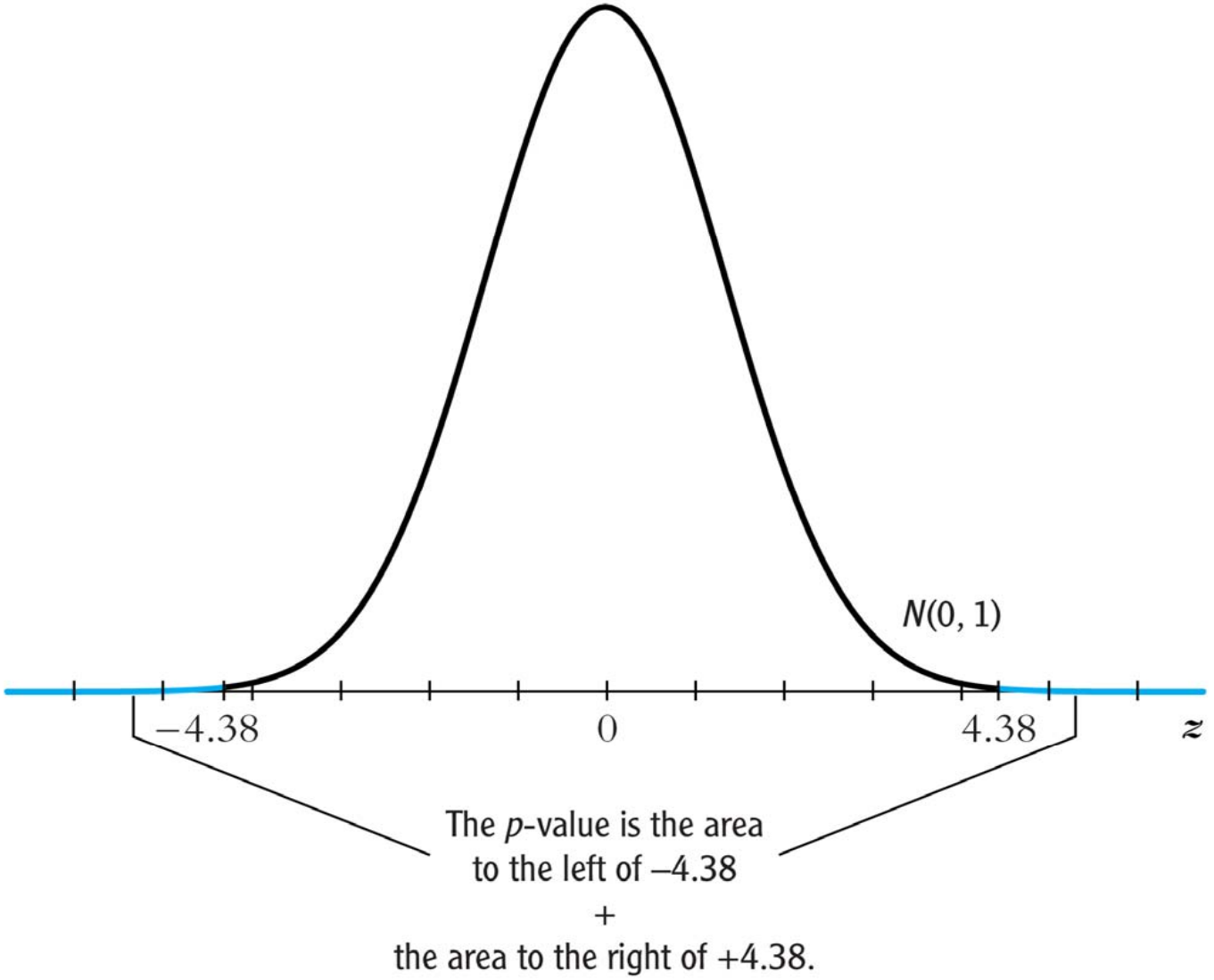
Estimated regression line: $\widehat{TestScore} = 698.9 - 2.28 \times STR$

Regression software reports the standard errors:

$$SE(\hat{\beta}_0) = 10.4 \qquad SE(\hat{\beta}_1) = 0.52$$

t-statistic testing $\beta_{1,0} = 0 = \dfrac{\hat{\beta}_1 - \beta_{1,0}}{SE(\hat{\beta}_1)} = \dfrac{-2.28 - 0}{0.52} = -4.38$

- The 1% 2-sided significance level is 2.58, so we reject the null at the 1% significance level.
- Alternatively, we can compute the p-value…



The *p*-value based on the large-*n* standard normal approximation to the *t*-statistic is 0.00001 (10^{-5})

Confidence Intervals for β_1

Recall that a 95% confidence is, equivalently:

- The set of points that cannot be rejected at the 5% significance level;
- A set-valued function of the data (an interval that is a function of the data) that contains the true parameter value 95% of the time in repeated samples.

Because the *t*-statistic for β_1 is N(0,1) in large samples, construction of a 95% confidence for β_1 is just like the case of the sample mean:

95% confidence interval for $\beta_1 = \{\hat{\beta}_1 \pm 1.96 \times SE(\hat{\beta}_1)\}$

Confidence interval example: *Test Scores* and *STR*

Estimated regression line: $\widehat{TestScore} = 698.9 - 2.28 \times STR$

$$SE(\hat{\beta}_0) = 10.4 \qquad SE(\hat{\beta}_1) = 0.52$$

95% confidence interval for $\hat{\beta}_1$:

$$\{\hat{\beta}_1 \pm 1.96 \times SE(\hat{\beta}_1)\} = \{-2.28 \pm 1.96 \times 0.52\}$$
$$= (-3.30, -1.26)$$

The following two statements are equivalent *(why?)*

- The 95% confidence interval does not include zero;
- The hypothesis $\beta_1 = 0$ is rejected at the 5% level

A concise (and conventional) way to report regressions:

Put standard errors in parentheses below the estimated coefficients to which they apply.

$$\widehat{TestScore} = 698.9 - 2.28 \times STR,\ R^2 = .05,\ SER = 18.6$$
$$(10.4)\ \ (0.52)$$

<u>This expression gives a lot of information</u>

- The estimated regression line is
$$\widehat{TestScore} = 698.9 - 2.28 \times STR$$
- The standard error of $\hat{\beta}_0$ is 10.4
- The standard error of $\hat{\beta}_1$ is 0.52
- The R^2 is .05; the standard error of the regression is 18.6

OLS regression: reading STATA output

```
regress testscr str, robust

Regression with robust standard errors                Number of obs =     420
                                                      F(  1,   418) =   19.26
                                                      Prob > F      =  0.0000
                                                      R-squared     =  0.0512
                                                      Root MSE      =  18.581
------------------------------------------------------------------------------
             |               Robust
    testscr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         str |  -2.279808   .5194892    -4.38   0.000    -3.300945   -1.258671
       _cons |    698.933   10.36436     67.44   0.000     678.5602    719.3057
------------------------------------------------------------------------------
```

so:

$$\widehat{TestScore} = 698.9 - 2.28 \times STR, \quad , \quad R^2 = .05, \; SER = 18.6$$

$$(10.4) \; (0.52)$$

$t\,(\beta_1 = 0) = -4.38$, p-value = **0.000** (2-sided)

95% 2-sided conf. interval for β_1 is (–3.30, –1.26)

Summary of Statistical Inference about β_0 and β_1:

Estimation:

- OLS estimators $\hat{\beta}_0$ and $\hat{\beta}_1$
- $\hat{\beta}_0$ and $\hat{\beta}_1$ have approximately normal sampling distributions in large samples

Testing:

- H_0: $\beta_1 = \beta_{1,0}$ v. $\beta_1 \neq \beta_{1,0}$ ($\beta_{1,0}$ is the value of β_1 under H_0)
- $t = (\hat{\beta}_1 - \beta_{1,0})/SE(\hat{\beta}_1)$
- p-value = area under standard normal outside t^{act} (large n)

Confidence Intervals:

- 95% confidence interval for β_1 is $\{\hat{\beta}_1 \pm 1.96 \times SE(\hat{\beta}_1)\}$
- This is the set of β_1 that is not rejected at the 5% level
- The 95% CI contains the true β_1 in 95% of all samples.

Regression when X is Binary

Sometimes a regressor is binary:

- $X = 1$ if small class size, = 0 if not
- $X = 1$ if female, = 0 if male
- $X = 1$ if treated (experimental drug), = 0 if not

Binary regressors are sometimes called "dummy" variables.

So far, β_1 has been called a "slope," but that doesn't make sense if X is binary.

How do we interpret regression with a binary regressor?

Interpreting regressions with a binary regressor

$Y_i = \beta_0 + \beta_1 X_i + u_i$, where X is binary ($X_i = 0$ or 1):

When $X_i = 0$, $Y_i = \beta_0 + u_i$

- the mean of Y_i is β_0
- that is, $E(Y_i|X_i=0) = \beta_0$

When $X_i = 1$, $Y_i = \beta_0 + \beta_1 + u_i$

- the mean of Y_i is $\beta_0 + \beta_1$
- that is, $E(Y_i|X_i=1) = \beta_0 + \beta_1$

so:

$$\beta_1 = E(Y_i|X_i=1) - E(Y_i|X_i=0)$$

= population difference in group means

Example: Let $D_i = \begin{cases} 1 \text{ if } STR_i \leq 20 \\ 0 \text{ if } STR_i > 20 \end{cases}$

OLS regression: $\widehat{TestScore} = \mathbf{650.0} + \mathbf{7.4} \times D$

(1.3) (**1.8**)

Tabulation of group means:

Class Size	**Average score ($\bar{Y}$)**	**Std. dev. (s_Y)**	***N***
Small ($STR \leq 20$)	657.4	19.4	238
Large ($STR > 20$)	**650.0**	17.9	182

Difference in means: $\bar{Y}_{\text{small}} - \bar{Y}_{\text{large}} = 657.4 - 650.0 = \mathbf{7.4}$

Standard error: $SE = \sqrt{\frac{s_s^2}{n_s} + \frac{s_l^2}{n_l}} = \sqrt{\frac{19.4^2}{238} + \frac{17.9^2}{182}} = \mathbf{1.8}$

Hypothesis testing

Difference-in-means test: compute the *t*-statistic,

$$t = \frac{\bar{Y}_s - \bar{Y}_l}{\sqrt{\frac{s_s^2}{n_s} + \frac{s_l^2}{n_l}}} = \frac{\bar{Y}_s - \bar{Y}_l}{SE(\bar{Y}_s - \bar{Y}_l)}$$

where $SE(\bar{Y}_s - \bar{Y}_l)$ is the "standard error" of $\bar{Y}_s - \bar{Y}_l$; the subscripts *s* and *l* refer to "small" and "large" STR districts; and $s_s^2 = \frac{1}{n_s - 1}\sum_{i=1}^{n_s}(Y_i - \bar{Y}_s)^2$ (etc.)

Compute the difference-of-means *t*-statistic:

Size	$\bar{Y}$	s_Y	n
small	657.4	19.4	238
large	650.0	17.9	182

$$t = \frac{\bar{Y}_s - \bar{Y}_l}{\sqrt{\frac{s_s^2}{n_s} + \frac{s_l^2}{n_l}}} = \frac{657.4 - 650.0}{\sqrt{\frac{19.4^2}{238} + \frac{17.9^2}{182}}} = \frac{7.4}{1.83} = 4.05$$

|t| > 1.96, so reject (at the 5% significance level) the null hypothesis that the two means are the same.

Confidence interval

A 95% confidence interval for the difference between the means is,

$$(\bar{Y}_s - \bar{Y}_l) \pm 1.96 \times SE(\bar{Y}_s - \bar{Y}_l)$$

$$= 7.4 \pm 1.96 \times 1.83 = (3.8, 11.0)$$

Two equivalent statements:

1. The 95% confidence interval for Δ doesn't include 0;
2. The hypothesis that $\Delta = 0$ is rejected at the 5% level.

Summary: regression when X_i is binary (0/1)

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

- β_0 = mean of Y when $X = 0$
- $\beta_0 + \beta_1$ = mean of Y when $X = 1$
- β_1 = difference in group means, X =1 minus $X = 0$
- SE($\hat{\beta}_1$) has the usual interpretation
- t-statistics, confidence intervals constructed as usual
- This is another way (an easy way) to do difference-in-means analysis
- The regression formulation is especially useful when we have additional regressors

Heteroskedasticity and Homoskedasticity, and Homoskedasticity-Only Standard Errors

- What…?
- Consequences of homoskedasticity
- Implication for computing standard errors

What do these two terms mean?

If var($u|X=x$) is constant – that is, if the variance of the conditional distribution of u given X does not depend on X – then u is said to be ***homoskedastic***. Otherwise, u is ***heteroskedastic***.

Example: hetero/homoskedasticity in the case of a binary regressor (that is, the comparison of means)

- Standard error when group variances are ***unequal***:

$$SE = \sqrt{\frac{s_s^2}{n_s} + \frac{s_l^2}{n_l}}$$

- Standard error when group variances are ***equal***:

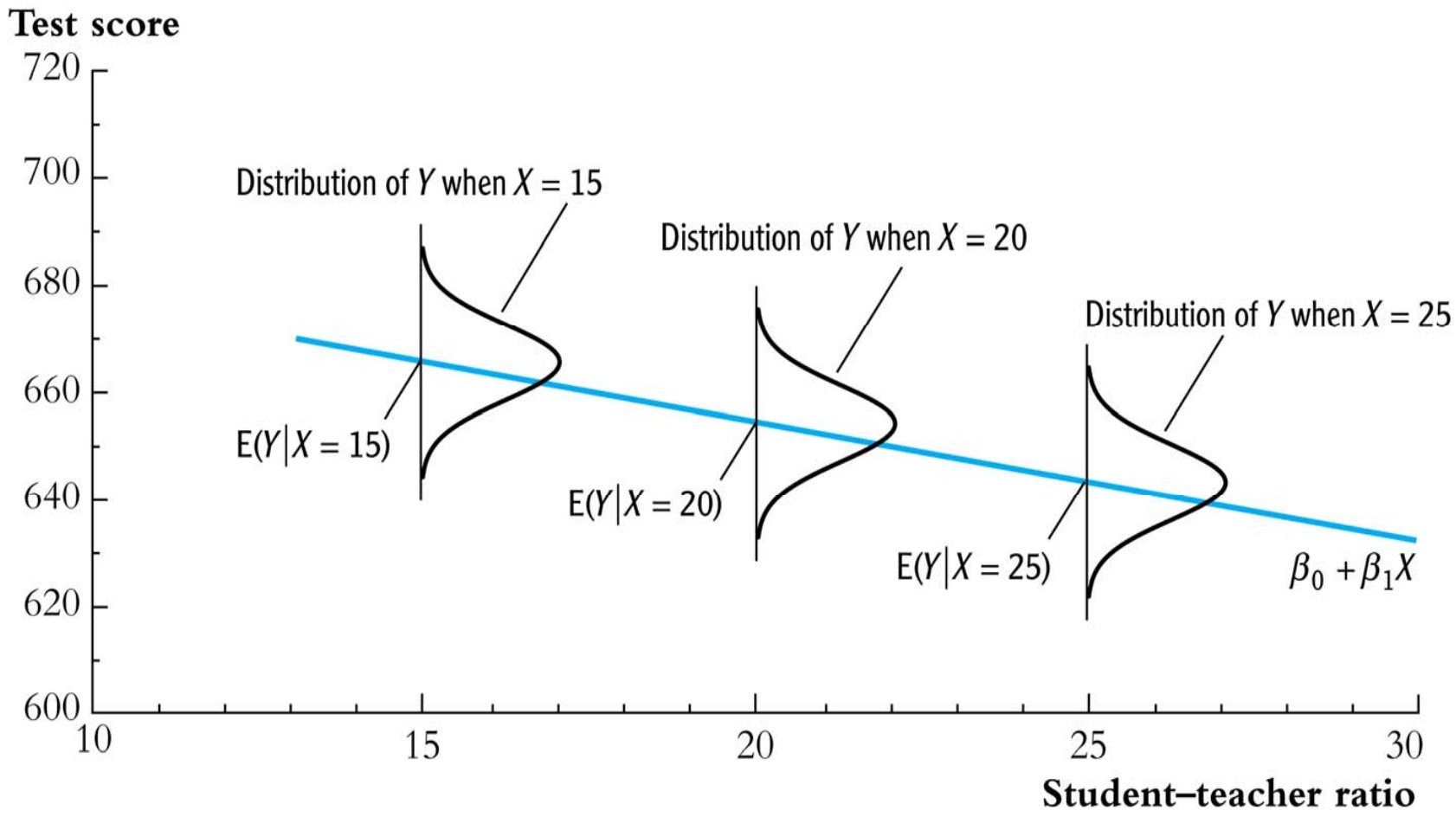
$$SE = s_p \sqrt{\frac{1}{n_s} + \frac{1}{n_l}}$$

where $s_p^2 = \dfrac{(n_s - 1)s_s^2 + (n_l - 1)s_l^2}{n_s + n_l - 2}$

s_p = "pooled estimator of σ^2" when $\sigma_l^2 = \sigma_s^2$

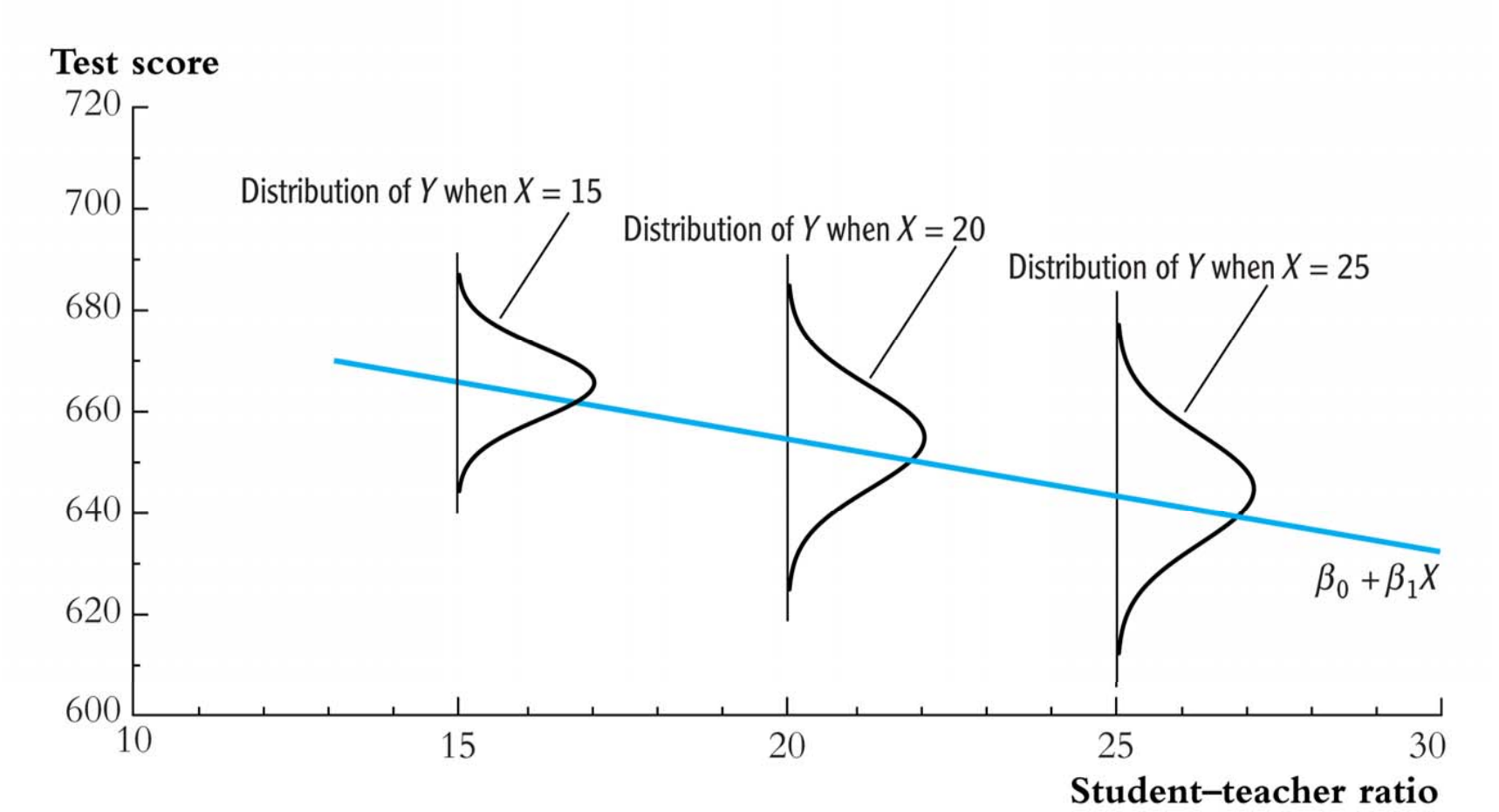
- **Equal** group variances = **homo**skedasticity
- **Unequal** group variances = **hetero**skedasticity

Homoskedasticity *in a picture:*



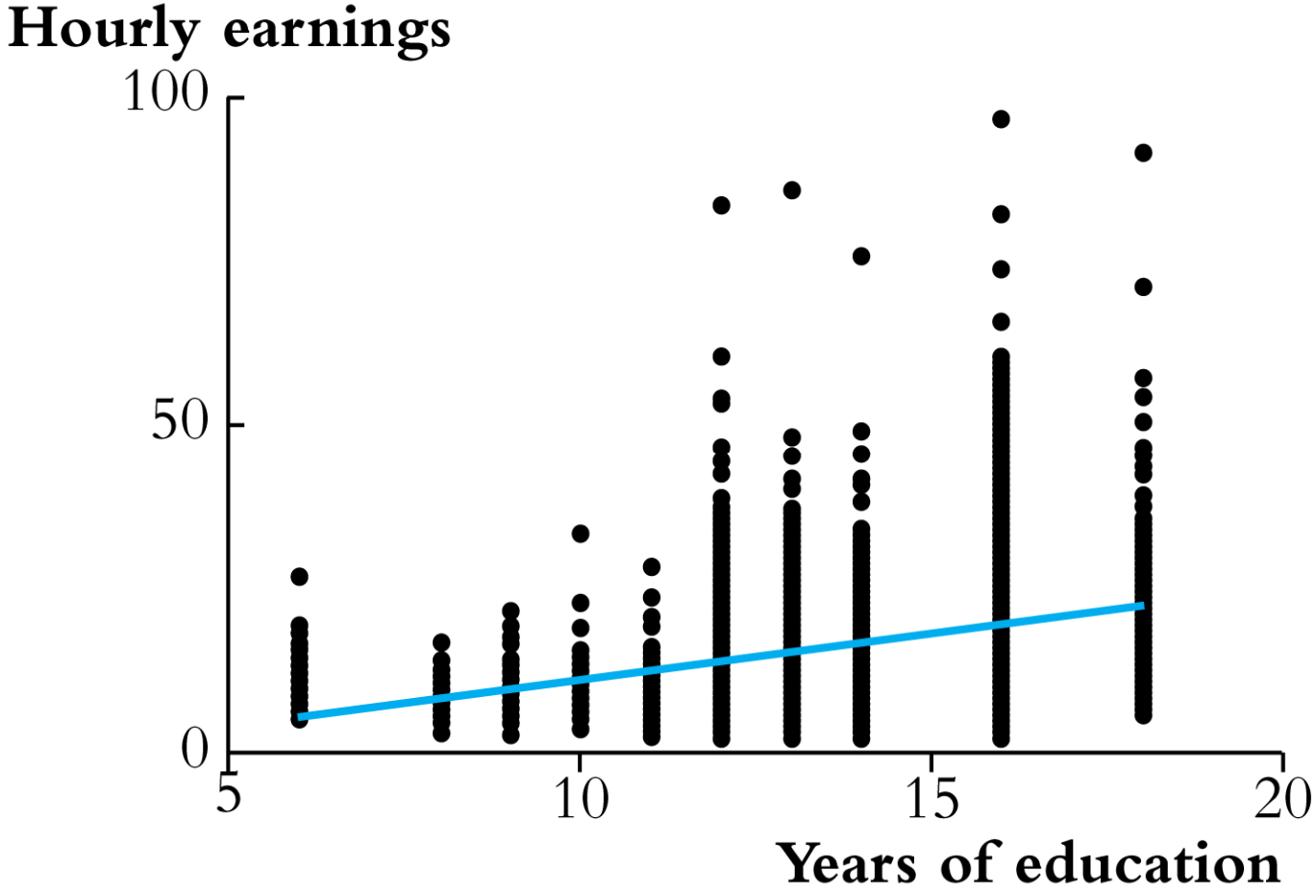
- $E(u|X=x) = 0$ (u satisfies Least Squares Assumption #1)
- The variance of u ***does not*** depend on x

Heteroskedasticity *in a picture:*



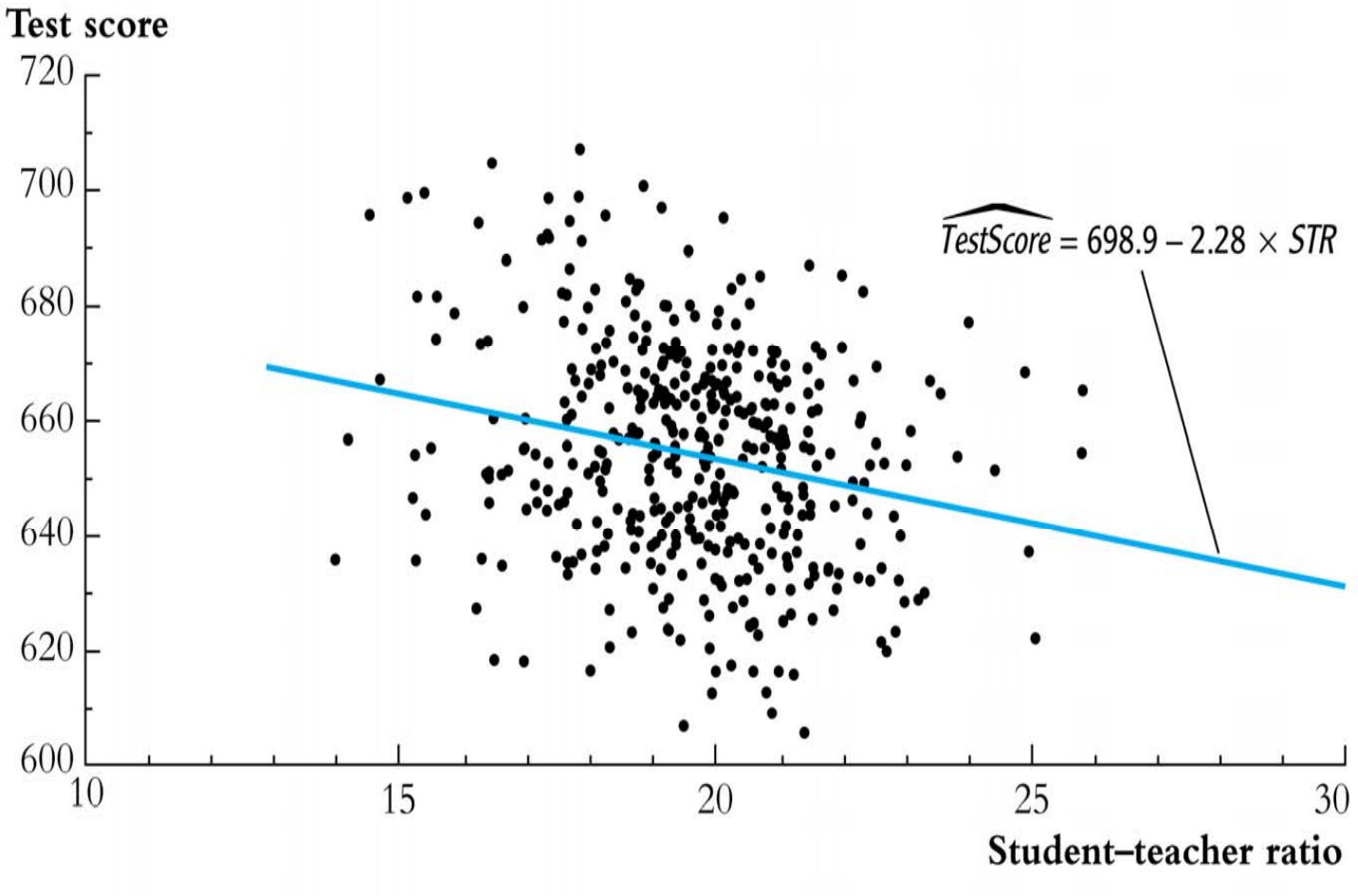
- $E(u|X=x) = 0$ (u satisfies Least Squares Assumption #1)
- The variance of u ***does*** depends on x: u is heteroskedastic.

A real-data example from labor economics: average hourly earnings vs. years of education (data source: Current Population Survey):



Heteroskedastic or homoskedastic?

The class size data:



Heteroskedastic or homoskedastic?

So far we have (without saying so) assumed that u might be heteroskedastic.

Recall the three least squares assumptions:

1. $E(u|X = x) = 0$
2. (X_i,Y_i), i =1,…,n, are i.i.d.
3. Large outliers are rare

Heteroskedasticity and homoskedasticity concern var($u|X=x$). Because we have not explicitly assumed homoskedastic errors, we have implicitly allowed for heteroskedasticity.

What if the errors are in fact homoskedastic?

- You can prove that OLS has the lowest variance among unbiased estimators that are linear in *Y*… a result called the Gauss-Markov theorem that we will return to shortly.
- The formula for the variance of $\hat{\beta}_1$ and the OLS standard error simplifies: If $\text{var}(u_i|X_i=x) = \sigma_u^2$, then

$$\text{var}(\hat{\beta}_1) = \frac{\text{var}[(X_i - \mu_x)u_i]}{n(\sigma_X^2)^2} = \frac{E[(X_i - \mu_x)^2 u_i^2]}{n(\sigma_X^2)^2}$$

$$= \frac{\sigma_u^2}{n\sigma_X^2}$$

Note: $\text{var}(\hat{\beta}_1)$ is inversely proportional to var(*X*): more spread in *X* means more information about $\hat{\beta}_1$ - we discussed this earlier but it is clearer from this formula.

- Along with this homoskedasticity-only formula for the variance of $\hat{\beta}_1$, we have homoskedasticity-only standard errors:

  Homoskedasticity-only standard error formula:

$$SE(\hat{\beta}_1) = \sqrt{\frac{1}{n} \times \frac{\frac{1}{n-2}\sum_{i=1}^{n} \hat{u}_i^2}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}}.$$

  Some people (e.g. Excel programmers) find the homoskedasticity-only formula simpler.

We now have two formulas for standard errors for $\hat{\beta}_1$.

- ***Homoskedasticity-only standard errors*** – these are valid only if the errors are homoskedastic.
- The usual standard errors – to differentiate the two, it is conventional to call these ***heteroskedasticity – robust standard errors***, because they are valid whether or not the errors are heteroskedastic.
- The main advantage of the homoskedasticity-only standard errors is that the formula is simpler. But the disadvantage is that the formula is only correct in general if the errors are homoskedastic.

Practical implications…

- The homoskedasticity-only formula for the standard error of $\hat{\beta}_1$ and the "heteroskedasticity-robust" formula differ – so in general, *you get different standard errors using the different formulas*.
- Homoskedasticity-only standard errors are the default setting in regression software – sometimes the only setting (e.g. Excel). To get the general "heteroskedasticity-robust" standard errors you must override the default.
- **If you don't override the default and there is in fact heteroskedasticity, your standard errors (and wrong *t*-statistics and confidence intervals) will be wrong – typically, homoskedasticity-only *SE*s are too small.**

Heteroskedasticity-robust standard errors in STATA

```
regress testscr str, robust

Regression with robust standard errors                Number of obs =     420
                                                      F(  1,   418) =   19.26
                                                      Prob > F      =  0.0000
                                                      R-squared     =  0.0512
                                                      Root MSE      =  18.581
------------------------------------------------------------------------------
             |               Robust
    testscr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         str |  -2.279808   .5194892    -4.39   0.000    -3.300945   -1.258671
       _cons |    698.933   10.36436    67.44   0.000     678.5602    719.3057
------------------------------------------------------------------------------
```

- If you use the "**, robust**" option, STATA computes heteroskedasticity-robust standard errors
- Otherwise, STATA computes homoskedasticity-only standard errors

The bottom line:

- If the errors are either homoskedastic or heteroskedastic and you use heteroskedastic-robust standard errors, you are OK
- If the errors are heteroskedastic and you use the homoskedasticity-only formula for standard errors, your standard errors will be wrong (the homoskedasticity-only estimator of the variance of $\hat{\beta}_1$ is inconsistent if there is heteroskedasticity).
- The two formulas coincide (when *n* is large) in the special case of homoskedasticity
- So, you should always use heteroskedasticity-robust standard errors.

Some Additional Theoretical Foundations of OLS

We have already learned a very great deal about OLS: OLS is unbiased and consistent; we have a formula for heteroskedasticity-robust standard errors; and we can construct confidence intervals and test statistics.

Also, a very good reason to use OLS is that everyone else does – so by using it, others will understand what you are doing. In effect, OLS is the language of regression analysis, and if you use a different estimator, you will be speaking a different language.

Still, some of you may have further questions:

- Is this really a good reason to use OLS? Aren't there other estimators that might be better – in particular, ones that might have a smaller variance?
- Also, what ever happened to our old friend, the Student *t* distribution?

So we will now answer these questions – but to do so we will need to make some stronger assumptions than the three least squares assumptions already presented.

The Extended Least Squares Assumptions

These consist of the three LS assumptions, plus two more:

1. $E(u|X = x) = 0$.
2. (X_i,Y_i), $i = 1,\ldots,n$, are i.i.d.
3. Large outliers are rare ($E(Y^4) < \infty$, $E(X^4) < \infty$).
4. u is homoskedastic
5. u is distributed $N(0,\sigma^2)$

- Assumptions 4 and 5 are more restrictive – so they apply to fewer cases in practice. However, if you make these assumptions, then certain mathematical calculations simplify and you can prove strong results – results that hold if these additional assumptions are true.
- We start with a discussion of the efficiency of OLS

Efficiency of OLS, part I: The Gauss-Markov Theorem

Under extended LS assumptions 1-4 (the basic three, plus homoskedasticity), $\hat{\beta}_1$ has the smallest variance among *all unbiased linear estimators* (unbiased estimators that are linear functions of $Y_1,\ldots, Y_n$). This is the ***Gauss-Markov theorem***.

Comments

- The GM theorem is proven in SW Appendix 5.2

The Gauss-Markov Theorem, ctd.

- $\hat{\beta}_1$ is a linear estimator, that is, it can be written as a linear function of $Y_1,\ldots, Y_n$:

$$\hat{\beta}_1 - \beta_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})u_i}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{1}{n}\sum_{i=1}^{n} w_i u_i ,$$

where $w_i = \dfrac{(X_i - \bar{X})}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}$.

- The G-M theorem says that among all possible choices of $\{w_i\}$, the OLS weights yield the smallest var($\hat{\beta}_1$)

Efficiency of OLS, part II:

- Under <u>all five</u> extended LS assumptions – including normally distributed errors – $\hat{\beta}_1$ has the smallest variance of *<u>all</u>* consistent estimators (linear *or* nonlinear functions of $Y_1,\ldots,Y_n$), as $n \to \infty$.
- This is a pretty amazing result – it says that, if (in addition to LSA 1-3) the errors are homoskedastic and normally distributed, then OLS is a better choice than any other consistent estimator. And because an estimator that isn't consistent is a poor choice, this says that OLS really is the best you can do – if all five extended LS assumptions hold. (The proof of this result is beyond the scope of this course and it is typically done in graduate courses.)

Some not-so-good thing about OLS

The foregoing results are impressive, but these results – and the OLS estimator – have important limitations.

1. The GM theorem really isn't that compelling:
   - The condition of homoskedasticity often doesn't hold (homoskedasticity is special)
   - The result is only for linear estimators – only a small subset of estimators (more on this in a moment)

2. The strongest optimality result ("part II" above) requires homoskedastic normal errors – not plausible in applications (think about the hourly earnings data!)

Limitations of OLS, ctd.

3. OLS is more sensitive to outliers than some other estimators. In the case of estimating the population mean, if there are big outliers, then the median is preferred to the mean because the median is less sensitive to outliers – it has a smaller variance than OLS when there are outliers. Similarly, in regression, OLS can be sensitive to outliers, and if there are big outliers other estimators can be more efficient (have a smaller variance). One such estimator is the least absolute deviations (LAD) estimator:

$$\min_{b_0, b_1} \sum_{i=1}^{n} \left| Y_i - (b_0 + b_1 X_i) \right|$$

In virtually all applied regression analysis, OLS is used – and that is what we will do in this course too.

Inference if *u* is Homoskedastic and Normal: the Student *t* Distribution (Section 5.6)

Recall the five extended LS assumptions:

1. $E(u|X = x) = 0$.
2. (X_i,Y_i), $i = 1,\ldots,n$, are i.i.d.
3. Large outliers are rare ($E(Y^4) < \infty$, $E(X^4) < \infty$).
4. u is homoskedastic
5. u is distributed $N(0,\sigma^2)$

If all five assumptions hold, then:

- $\hat{\beta}_0$ and $\hat{\beta}_1$ are normally distributed *for all n* (!)
- the t-statistic has a Student t distribution with $n - 2$ degrees of freedom – this holds exactly *for all n* (!)

Normality of the sampling distribution of $\hat{\beta}_1$ under 1–5:

$$\hat{\beta}_1 - \beta_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})u_i}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

$$= \frac{1}{n}\sum_{i=1}^{n} w_i u_i \text{, where } w_i = \frac{(X_i - \bar{X})}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}.$$

What is the distribution of a weighted average of normals?

Under assumptions 1 – 5:

$$\hat{\beta}_1 - \beta_1 \sim N\left(0, \frac{1}{n^2}\left(\sum_{i=1}^{n} w_i^2\right)\sigma_u^2\right) \qquad (*)$$

Substituting w_i into (*) yields the homoskedasticity-only variance formula.

In addition, under assumptions 1 – 5, under the null hypothesis the t statistic has a Student t distribution with $n - 2$ degrees of freedom

- Why $n - 2$? because we estimated 2 parameters, β_0 and β_1
- For $n < 30$, the t critical values can be a fair bit larger than the $N(0,1)$ critical values
- For $n > 50$ or so, the difference in t_{n-2} and $N(0,1)$ distributions is negligible. Recall the Student t table:

degrees of freedom	5% t-distribution critical value
10	2.23
20	2.09
30	2.04
60	2.00
∞	1.96

Practical implication:

- If $n < 50$ ***and*** you really believe that, for your application, u is homoskedastic and normally distributed, then use the t_{n-2} instead of the $N(0,1)$ critical values for hypothesis tests and confidence intervals.
- In most econometric applications, there is no reason to believe that u is homoskedastic and normal – usually, there is good reason to believe that neither assumption holds.
- Fortunately, in modern applications, $n > 50$, so we can rely on the large-n results presented earlier, based on the CLT, to perform hypothesis tests and construct confidence intervals using the large-n normal approximation.

Summary and Assessment (Section 5.7)

- The initial policy question:

  Suppose new teachers are hired so the student-teacher ratio falls by one student per class. What is the effect of this policy intervention ("treatment") on test scores?

- Does our regression analysis answer this convincingly?
  - *Not really* – districts with low *STR* tend to be ones with lots of other resources and higher income families, which provide kids with more learning opportunities outside school…this suggests that corr(u_i, STR_i) > 0, so $E(u_i|X_i) \neq 0$.
  - So, we have omitted some factors, or variables, from our analysis, and this has biased our results.