## **Applied Econometrics**

Multiple Regression Analysis

Interval Estimation and Hypothesis Testing I

Text: Wooldridge, Chapter 4 and Appendix C

June 30, 2011

# Sampling Distribution of the OLS Estimator under Normality

• Consider the OLS estimator  $\widehat{oldsymbol{eta}}$  of  $oldsymbol{eta}$  in the model

$$y = X\beta + u, \tag{1}$$

i.e.,

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}.$$
(2)

• We know that

$$\mathsf{E}(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta} \tag{3}$$

$$\operatorname{Cov}(\widehat{\boldsymbol{\beta}}) = \sigma^2 (\boldsymbol{X}' \boldsymbol{X})^{-1}.$$
 (4)

• In order to conduct statistical inference such as hypothesis testing, we need to be able to make statements about the full sampling distribution of  $\hat{\beta}$ .

#### **The Normality Assumption**

- To derive the distributions of relevant test statistics, we add to our earlier *GauB–Markov Assumptions* the assumption of normality:
- Normality assumption: The error u is independent of X and normally distributed with mean zero and variance  $\sigma^2$ , i.e.,

$$u_i \sim \mathsf{N}(0, \sigma^2). \tag{5}$$

• Given our earlier result that  $Cov(u) = \sigma^2 I$ , the normality assumption can also be written as

$$\boldsymbol{u} \sim \mathsf{MVN}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}),$$
 (6)

indicating a multivariate normal (MVN) distribution with mean 0 and covariance matrix  $\sigma^2 I$ .

• With the assumption of normality added, the linear regression model under the Gauß–Markov Assumptions is also referred to as the **classical linear model (CLM)**.

#### **Implications of the Normality Assumption**

• Recall that

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u})$$
  
=  $\boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u},$ 

so that  $\widehat{oldsymbol{eta}}$  is a linear function of the normally distributed  $oldsymbol{u}.$ 

- In general, if 
$$oldsymbol{x} \sim \mathsf{MVN}(oldsymbol{\mu}, oldsymbol{\Sigma})$$
, then

$$Ax + b \sim \mathsf{MVN}(A\mu + b, A\Sigma A').$$
 (7)

• Hence  $\widehat{\beta}$  is likewise normal, i.e.,

$$\widehat{\boldsymbol{\beta}}|\boldsymbol{X} \sim \mathsf{MVN}(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}),$$
 (8)

i.e., the OLS estimator  $\hat{\beta}$  is normally distributed with mean vector  $\beta$  and covariance matrix  $\sigma^2 (X'X)^{-1}$ .

#### **Implications of the Normality Assumption**

- Implications of (8), among others, are as follows:
- (1) Each element of  $\hat{\beta}$ , say  $\hat{\beta}_j$ , has a univariate normal distribution with mean  $\beta_j$  and a variance given by the corresponding *diagonal element* of the covariance matrix  $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ , i.e.,

$$\widehat{\beta}_j \sim \mathsf{N}(\beta_j, \sigma_{\widehat{\beta}_j}^2),$$
(9)

where  $\sigma_{\widehat{\beta}_j}^2$  denotes the sampling variance of  $\widehat{\beta}_j$ , and by standardization we get the standard normal with mean zero and unit variance,

$$\frac{\widehat{\beta}_j - \beta_j}{\sigma_{\widehat{\beta}_j}} \sim \mathsf{N}(0, 1), \tag{10}$$

where  $\sigma_{\widehat{\beta}_j} = \sqrt{\sigma_{\widehat{\beta}_j}^2}$  is the standard error of  $\widehat{\beta}_j$ .

• For example, in the simple linear regression model,  $y = \beta_0 + \beta_1 x_1 + u$ ,

$$\sigma^2 (\mathbf{X}' \mathbf{X})^{-1} = \frac{\sigma^2}{n s_{x_1}^2} \begin{bmatrix} \overline{x_1^2} & -\overline{x}_1 \\ -\overline{x}_1 & 1 \end{bmatrix},$$

SO

$$\widehat{\beta}_0 \sim \mathsf{N}\left(\beta_0, \frac{\sigma^2 \overline{x_1^2}}{n s_{x_1}^2}\right),$$

 $\mathsf{and}$ 

$$\widehat{\beta}_1 \sim \mathsf{N}\left(\beta_1, \frac{\sigma^2}{ns_{x_1}^2}\right).$$

- A further implication of (8) is:
- (2) All linear combinations of the elements in  $\hat{\beta}$  are likewise normally distributed.
  - That is, for matrix  $\boldsymbol{R}$ , it follows from (7) that

$$\mathbf{R}\widehat{\boldsymbol{\beta}} \sim \mathsf{MVN}\left(\mathbf{R}\boldsymbol{\beta}, \sigma^{2}\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\right).$$
 (11)

## **Interval Estimation and Confidence Intervals**

- The OLS estimator  $\hat{\beta}_j$  is a point estimator of  $\beta_j$ .
- With probability one,  $\widehat{\beta}_j \neq \beta_j$ .
- For practical applications, we are also interested in knowing a likely range of values for the unknown parameter, given the data.
- That is, we want to construct an interval estimator for  $\beta_j$  (called a **confidence interval, CI**), which *a priori* has a high probability of containing the unknown parameter.
- Such an interval also reflects the precision of our point estimate.
- We construct the confidence interval in such a way that it contains the true parameter value β<sub>j</sub> with some pre-specified probability (coverage probability or confidence coefficient) 1 α, where α is the (typically small) probability of error.
- E.g., if  $\alpha = 0.05$ , then the CI will contain  $\beta_j$  with probability 0.95.

• Formal definition:  $CI_{1-\alpha}$  is a *random interval* such that, conditional on X,

$$\Pr(\beta_j \in \mathsf{CI}_{1-\alpha}) = 1 - \alpha. \tag{12}$$

- Note that the random element in (12) is  $CI_{1-\alpha}$ ,  $\beta_j$  is a fixed quantity (the true population parameter).
- We consider a symmetric confidence interval around the estimator  $\widehat{\beta}_j$ :

$$\mathsf{CI}_{1-\alpha} = [\widehat{\beta}_j - c_{1-\alpha}, \widehat{\beta}_j + c_{1-\alpha}].$$
(13)

• We need to determine  $c_{1-\alpha}$  such that

$$\Pr(\beta_j \in \mathsf{CI}_{1-\alpha}) = 1 - \alpha. \tag{14}$$

• We know that

$$\widehat{\beta}_j \sim \mathsf{N}(\beta_j, \sigma_{\widehat{\beta}_j}^2), \quad \sigma_{\widehat{\beta}_j}^2 = \sigma^2 (X'X)_{jj}^{-1},$$
(15)

where  $(X'X)_{jj}^{-1}$  is the corresponding diagonal element of  $(X'X)^{-1}$ , and

$$\frac{\widehat{\beta}_j - \beta_j}{\sqrt{\sigma_{\widehat{\beta}_j}^2}} \sim \mathsf{N}(0, 1).$$
(16)

- Let  $Z \sim N(0,1)$  (i.e., Z has a standard normal distributon).
- The  $\gamma$ -quantile of the standard normal distribution  $z_{\gamma}$  is defined such that

$$\Pr(Z \le z_{\gamma}) = \int_{-\infty}^{z_{\gamma}} \frac{e^{-z^2/2} dz}{\sqrt{2\pi}} = \gamma.$$
(17)

• By symmetry around zero,

$$z_{1-\gamma} = -z_{\gamma}. \tag{18}$$

• For example, if  $\gamma = 0.975$ , then  $z_{\gamma} = 1.96$ , and so  $z_{0.025} = -1.96$ .

Table 1: Quantiles of the standard normal distribution

 $\gamma$	0.9	0.95	0.975	0.99	0.995
$z_\gamma$	1.282	1.645	1.960	2.326	2.576

- Now suppose for the moment that  $\sigma^2$  and hence  $\sigma^2_{\widehat{\beta}_j}$  is known.
- Take  $\gamma = 1 \alpha/2$ , and consider the interval

$$\mathsf{CI} = \left[\widehat{\beta}_j - z_{1-\alpha/2} \times \sigma_{\widehat{\beta}_j}, \widehat{\beta}_j + z_{1-\alpha/2} \times \sigma_{\widehat{\beta}_j}\right]. \tag{19}$$

• Then

$$\Pr(\beta_{j} \in \mathsf{CI}) = \Pr\left(\widehat{\beta}_{j} - z_{1-\alpha/2} \times \sigma_{\widehat{\beta}_{j}} \leq \beta_{j} \leq \widehat{\beta}_{j} + z_{1-\alpha/2} \times \sigma_{\widehat{\beta}_{j}}\right)$$

$$= \Pr\left(-z_{1-\alpha/2} \times \sigma_{\widehat{\beta}_{j}} \leq \beta_{j} - \widehat{\beta}_{j} \leq z_{1-\alpha/2} \times \sigma_{\widehat{\beta}_{j}}\right)$$

$$= \Pr\left(-z_{1-\alpha/2} \times \sigma_{\widehat{\beta}_{j}} \leq \widehat{\beta}_{j} - \beta_{j} \leq z_{1-\alpha/2} \times \sigma_{\widehat{\beta}_{j}}\right)$$

$$= \Pr\left(-z_{1-\alpha/2} \leq \frac{\widehat{\beta}_{j} - \beta_{j}}{\sigma_{\widehat{\beta}_{j}}} \leq z_{1-\alpha/2}\right)$$

$$= \Pr\left(\frac{\widehat{\beta}_{j} - \beta_{j}}{\sigma_{\widehat{\beta}_{j}}} \leq z_{1-\alpha/2}\right) - \Pr\left(\frac{\widehat{\beta}_{j} - \beta_{j}}{\sigma_{\widehat{\beta}_{j}}} \leq -z_{1-\alpha/2}\right)$$

$$= \frac{\Pr\left(\frac{\widehat{\beta}_{j} - \beta_{j}}{\sigma_{\widehat{\beta}_{j}}} \leq z_{1-\alpha/2}\right)}{1-\alpha/2} - \frac{\Pr\left(\frac{\widehat{\beta}_{j} - \beta_{j}}{\sigma_{\widehat{\beta}_{j}}} \leq z_{\alpha/2}\right)}{z_{\alpha/2}}$$

$$= 1 - \alpha/2 - \alpha/2 = 1 - \alpha,$$
(24)

where in (23)  $-z_{1-\alpha/2} = z_{\alpha/2}$  was used.

• Since  $\sigma^2$  is not known, it has to be estimated from the data via

$$\widehat{\sigma}^2 = \frac{1}{n-k-1} \sum_{i=1}^n \widehat{u}_i^2 = \frac{\widehat{u}'\widehat{u}}{n-k-1}.$$
(25)

• However, in this case, quantity

$$t_{\widehat{\beta}_j} = \frac{\widehat{\beta}_j - \beta_j}{\widehat{\sigma}_{\widehat{\beta}_j}}, \quad \widehat{\sigma}_{\widehat{\beta}_j}^2 = \widehat{\sigma}^2 (X'X)_{jj}^{-1}, \tag{26}$$

is no longer normally distributed.

• It can be shown that (26) has a Student's t distribution with n - k - 1 degrees of freedom, i.e.,

$$t_{\widehat{\beta}_j} = \frac{\widehat{\beta}_j - \beta_j}{\widehat{\sigma}_{\widehat{\beta}_j}} \sim t(n - k - 1).$$
(27)



- The t distribution has more probability mass in the tails than the normal.
- Hence, the quantiles are larger (in magnitude) than the corresponding quantiles of the normal distribution, and hence confidence intervals are wider.
- This reflects the larger uncertainty due to the additional sampling error embodied in the estimated error term variance.
- For  $\nu \to \infty$ , the t distribution approaches the normal distribution, where  $\nu$  denotes degrees of freedom.
- Often normal quantiles are used for  $\nu > 30$ .
- Several quantiles of the t distribution are provided in Table 2.

ν	0.9	0.95	0.975	0.99	0.995
1	3.078	6.314	12.706	31.821	63.657
3	1.638	2.353	3.182	4.541	5.841
5	1.476	2.015	2.571	3.365	4.032
10	1.372	1.812	2.228	2.764	3.169
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
25	1.316	1.708	2.060	2.485	2.787
30	1.310	1.697	2.042	2.457	2.750
50	1.299	1.676	2.009	2.403	2.678
100	1.290	1.660	1.984	2.364	2.626
$\infty$	1.282	1.645	1.960	2.326	2.576

Table 2: Quantiles of the t distribution ( $\nu$  denotes degrees of freedom)

 $\nu = \infty$  corresponds to the normal distribution.

• Let  $t_{\gamma}^{\nu}$  be the  $\gamma$ -quantile of the t distribution with  $\nu$  degrees of freedom.

• That is, for 
$$T \sim t(\nu)$$
,  

$$\Pr(T \leq t_{\gamma}^{\nu}) = \gamma.$$
(28)

• Just as the standard normal distribution, the *t* distribution is symmetric around zero, so

$$t_{1-\gamma}^{\nu} = -t_{\gamma}^{\nu}.\tag{29}$$

• Hence, a feasible confidence interval (i.e., with  $\sigma^2$  unknown) with confidence level  $1 - \alpha$  is

$$\mathsf{CI}_{1-\alpha} = [\widehat{\beta}_j - t_{1-\alpha/2}^{n-k-1} \times \widehat{\sigma}_{\widehat{\beta}_j}, \widehat{\beta}_j + t_{1-\alpha/2}^{n-k-1} \times \widehat{\sigma}_{\widehat{\beta}_j}]$$
(30)

A rule of thumb for reasonably large sample sizes (so t ≈ normal) is to calculate a 95% confidence interval for β<sub>j</sub> as β<sub>j</sub> ± 2 × σ<sub>β<sub>j</sub></sub>.

#### **Interpretation of Confidence Intervals**

- The confidence interval is a function of the data and thus is random.
- Before the sample of data is generated, the probability is  $1 \alpha$  that the data will be such that  $CI_{1-\alpha}$  contains the true value  $\beta_j$ .
- Once we have observed the data, we calculate the realization of the interval for our particular sample of data.
- This realization of  $CI_{1-\alpha}$  is no longer random, so it makes no sense to talk about the probability that the realized interval contains  $\beta_j$ .
- The realized interval either contains the true value  $\beta_j$  or not; no probability is involved.

## **Hypothesis Testing**

- A hypothesis is a statement about one or more unknown population parameters.
- Based on the data, we want to decide whether the statement is true or false.
- Typically, we have two competing hypotheses:
  - Null hypothesis (H<sub>0</sub>): The null hypothesis is presumed to be true until the data strongly suggest otherwise.
  - Alternative hypothesis (H<sub>1</sub>): A hypothesis against which the null is tested. It is presumed to be true if the data reject the null.
- Usually the case of interest is stated as  $H_1$ , which can be accepted if the data provide strong evidence against  $H_0$ .
- We cannot "accept the null hypothesis"; either the evidence in the data rejects  $H_0$  or "fails to reject  $H_0$ " (see below for explanation).

• Consider

$$wage = \beta_0 + \beta_1 \times female + \beta_2 \times education + \dots + u.$$
 (31)

For example, we may test  $H_0: \beta_1 = 0$  (no wage discrimination) against  $H_1: \beta_1 < 0$  (wage discrimination).

- In hypothesis testing, there are two types of errors one can make:
  - Type I error: Rejecting  $H_0$  when  $H_0$  is true
  - The probability of a Type I error is denoted by  $\alpha$  and called the **significance level** of the test, i.e.,

$$\alpha = \Pr(\text{Reject } H_0 | H_0 \text{ is true}). \tag{32}$$

- Type II error: Not rejecting  $H_0$  when  $H_1$  is true.
- In hypothesis testing, we initially specify a significance level  $\alpha$  of a test.
- This is typically chosen as a small number such as 0.01, 0.05, or 0.1, so that a rejection of  $H_0$  can actually be viewed as strong evidence against the null (and in favor of  $H_1$ ).

- We have to determine a **test statistic** (or simply a **statistic**), and a **critical region** (or **rejection region**).
- The test statistic is a function of the data, i.e., a random quantity.
- Our particular sample of data gives rise to a particular realization of the test statistic.
- We then check whether the realization of the test statistic falls into the critical region, in which case we reject the null hypothesis.
- In summary, we have the following steps:
  - 1. Specify  $H_0$  and  $H_1$ .
  - 2. Choose the significance level  $\alpha$ .
  - 3. Define the critical region. This requires knowledge of the distribution of the test statistic *under the null hypothesis*.
  - 4. Calculate the test statistic for the observed sample of data and reject  $H_0$  if it falls into the critical region. Otherwise don't reject  $H_0$ .

#### Hypotheses About a Single Element of $\beta$ : The *t* Test

• Consider testing

$$H_0: \beta_j = \beta_j^0 \tag{33}$$

against

$$H_1: \beta_j \neq \beta_j^0 \tag{34}$$

where  $\beta_j^0$  is the value of the population parameter  $\beta_j$  specified by the null hypothesis.

- Such a test is called a two-sided test because the alternative hypothesis *H*<sub>1</sub> allows for deviations in both directions of β<sub>1</sub> from the value specified by the null.
- Our test statistic is

$$t_{\widehat{\beta}_j} = \frac{\widehat{\beta}_j - \beta_j^0}{\widehat{\sigma}_{\widehat{\beta}_j}},$$

and we know that, if the null hypothesis holds, i.e.,  $\beta_j = \beta_j^0$ 

$$t_{\widehat{\beta}_j} \sim t(n-k-1). \tag{35}$$

- We reject the null hypothesis if the observed value of  $t_{\hat{\beta}_j}$  is rather unlikely under the null hypothesis, i.e., if  $t_{\hat{\beta}_i}$  is large in magnitude.
- More precisely, we reject  $H_0$  if, under the null hypothesis, the probability of observing a value of  $|t_{\hat{\beta}_j}|$  or larger is smaller than the prespecified **significance level**  $\alpha$ .
- The significance level  $\alpha$  implies a specific **critical value**, c, which is implicitly defined by

$$\Pr(|t_{\widehat{\beta}_j}| > c | \mathsf{H}_0 \text{ is true}) = \alpha, \tag{36}$$

i.e., c is the  $1-\alpha/2$  quantile of the t(n-k-1) distribution.

• We reject  $H_0$  if  $|t_{\widehat{\beta}_i}| > c = t_{1-\alpha/2}^{n-k-1}$ .

• Suppose that n - k - 1 = 18, and we want to test

$$H_0:\beta_j=\beta_j^0$$

against the the alternative hypothesis

$$H_1:\beta_j\neq\beta_j^0$$

at a significance level of 5%, i.e.,  $\alpha = 0.05$  (which is often used).

- By inspection of Table 2, we observe that the  $1 \alpha/2 = 1 0.05/2 = 0.975$  quantile of the t(18) distribution is 2.101.
- So we reject  $H_0$  at the 5% level if

$$|t_{\widehat{\beta}_j}| > 2.101 = c.$$

• The situation is illustrated in the following figure.



## Note on Terminology: "Accept $H_0$ " vs. "Fail to reject $H_0$ "

- When  $H_0$  is not rejected, this is reported as "we fail to reject  $H_0$  at the  $(100\alpha)\%$  level" rather than as " $H_0$  is accepted at the  $(100\alpha)\%$  level".
- The reason is that there are generally infinitely many hypotheses that cannot be rejected.
- In particular, we cannot reject  $H_0$ :  $\beta_j = \beta_j^0$  at significance level  $\alpha$  against a two-sided alternative if and only if  $\beta_j^0$  is contained in a  $1 \alpha$  confidence interval.
- To see this, note that we do not reject  $H_0$  if and only if

$$-t_{1-\alpha/2}^{n-k-1} \le \frac{\widehat{\beta}_j - \beta_j^0}{\widehat{\sigma}_{\widehat{\beta}_j}} \le t_{1-\alpha/2}^{n-k-1}$$
(37)

$$\Leftrightarrow \quad \widehat{\beta}_j - t_{1-\alpha/2}^{n-k-1} \widehat{\sigma}_{\widehat{\beta}_j} \le \beta_j^0 \le \widehat{\beta}_j + t_{1-\alpha/2}^{n-k-1} \widehat{\sigma}_{\widehat{\beta}_j}, \tag{38}$$

i.e.,  $\beta_j^0 \in \mathsf{Cl}_{1-\alpha}$ .

- Thus, for any  $\beta_j^0$  in the  $1 \alpha$  confidence interval, we cannot reject  $H_0: \beta_j = \beta_j^0$ .
- Note how knowing an  $1 \alpha$  confidence interval allows to test any hypothesis  $H_0: \beta_j = \beta_j^0$  against  $H_1: \beta_j \neq \beta_j^0$  at significance level  $\alpha$ .
- Failing to reject  $H_0$  simply means that the evidence in the data against  $H_0$  is not strong enough at the pre-specified significance level.
- But we cannot say we accept  $H_0$ .

• In applications, we are often interested in testing the hypothesis

$$H_0: \beta_j = 0, \tag{39}$$

where  $\beta_j^0$  in (33).

- Recall that  $\beta_j$  measures the partial effect of  $x_j$  on the expected value of y, after controlling for the other regressors.
- Thus, hypothesis (39) means that  $x_j$  has no partial effect on y.
- The alternative in this case is

$$H_1: \beta_j \neq 0. \tag{40}$$

If (39) is rejected against (40) at the, e.g., 5% level, then this is often expressed as "x<sub>j</sub> is statistically significant at the 5% level"; otherwise "x<sub>j</sub> is statistically insignificant at the 5% level".

### **Example: Wage Equation**

• Consider the wage equation

 $\log(wage) = \beta_0 + \beta_1 \cdot educ + \beta_2 \cdot exper + \beta_3 \cdot tenure + u, \quad (41)$ 

where

- wage is average hourly earnings,
- educ is years of eduction,
- exper is years of potential experience,
- *tenure* is years with current employer.
- We might be interested in testing whether the experience has no impact on wage after controlling for education and tenure, i.e., we specify

$$H_0: \beta_2 = 0.$$
 (42)

• We estimate

$$\widehat{\log(wage)} = \underbrace{0.2844}_{(0.1042)} + \underbrace{0.0920}_{(0.0073)} educ + \underbrace{0.0041}_{(0.0017)} exper + \underbrace{0.0221}_{(0.0031)} tenure,$$

where standard errors are given in parentheses.

- In this case, n = 526, k + 1 = 4, so n k 1 is rather large, and we can use the critical values implied by the normal distribution.
- Thus, at the 5% and 1% levels, we would have

$$c_{0.05} = 1.96$$
, and  $c_{0.01} = 2.576$ ,

respectively.

• The t statistic is

$$t_{\widehat{\beta}_2} = \frac{0.0041}{0.0017} = 2.412. \tag{43}$$

• Thus, we reject  $H_0$  at the 5% level, but not at the 1% level.

## The *p*-value

- This illustrates that there might be a certain degree of subjectivity in the approach, since the significance level is chosen ahead of time.
- As different people may prefer different significance levels, it is often useful to report the *p*-value of a test.
- The decision to reject or not reject  $H_0$  depends on the critical value  $t_{1-\alpha/2}^{n-k-1}$ .
- If the significance level  $\alpha$  becomes smaller, then we move farther into the tails to determine the critical value.
- That is, the smaller is  $\alpha$ , the more "difficult" it is to reject  $H_0$ .
- The *p*-value is the *smallest significance level at which the null hypothesis would be rejected*, given the observed value of the *t* statistic.
- Therefore, the *p*-value might be viewed as a measure for the strength of evidence against the null hypothesis: The smaller the value, the stronger the evidence against H<sub>0</sub>.

• When the p--value is reported, decisions can be made regarding any significance level  $\alpha$  according to the rule

$$p - \mathsf{value} \left\{ \begin{array}{l} \leq \\ > \end{array} \right\} \alpha \Leftrightarrow \begin{cases} \mathsf{reject} \ H_0 \\ \mathsf{do not reject} \ H_0. \end{cases}$$

- To find the *p*-value for our testing problem above, we have to find the probability that, under the null hypothesis,  $|t_{\widehat{\beta}_2}| > 2.412$ .
- To do so, we calculate, using the cumulative distribution function (cdf) of the standard normal distribution,

$$\begin{aligned} \Pr(t_{\widehat{\beta}_2} < -2.412 | H_0) + \Pr(t_{\widehat{\beta}_2} > 2.412 | H_0) &= 2\Pr(t_{\widehat{\beta}_2} < -2.412 | H_0) \\ &= 2\Pr(t_{\widehat{\beta}_2} > 2.412 | H_0) \\ &= 0.0159 = p - \text{value.} \end{aligned}$$

• Thus, the evidence against  $H_0$  appears to be strong.

## **Statistical Significance vs. Economic Significance**

- It might be argued, however, that the effect of experience is not that large **economically**.
- For example, increasing experience by 5 years increases (predicted)  $\log(wage)$  by  $5 \times 0.0041 = 0.0205$ , i.e., (predicted) wage is 2% higher.
- It is essential that the economic or practical implications of estimated parameters are always discussed, i.e., that their magnitude is taken into account in addition to their statistical significance, since economic significance need not always accompany statistical significance.
- This is related to the too-large sample size problem:<sup>1</sup> The t-statistic for H<sub>0</sub>: β<sub>j</sub> = 0 is

$$t_{\widehat{\beta}_j} = \frac{\widehat{\beta}_j}{\widehat{\sigma}_{\widehat{\beta}_j}}.$$

<sup>1</sup>See, e.g., Peter Kennedy: *A Guide to Econometrics*, Ch. 4, for discussion.

- Now almost any relevant independent variable will have *some* influence, however small, i.e.,  $\beta_j \neq 0$ , although the magnitude may be insignificant economically.
- As the sample size n increases,  $\hat{\sigma}_{\hat{\beta}_j}$  decreases, and so eventually the t statistic will eventually become statistically significant.
- Clearly this does not imply that we should use smaller samples—a larger sample means more information and more precise estimates, which is good.
- However, it should be discussed whether an effect of primary interest is economically meaningful.
- Such discussion may also include the uncertainty about the parameter, which can be fostered, e.g., by reporting confidence intervals.
- It seems also reasonable to use smaller significance levels as the sample size increases, but clearly this does not settle the issue of economic importance.

#### **One-sided Alternatives**

• On the basis of economic theory or common sense, we can often specify a **one-sided alternative** of the form

$$H_1: \beta_j > \beta_j^0. \tag{44}$$

• Formally, the null hypothesis is then

$$H_0: \beta_j \le \beta_j^0. \tag{45}$$

- In statistical terms, the null hypothesis (45) is *composite*, i.e., it contains more than a single value for  $\beta_i$ .
- The probability of rejection depends on  $\beta_j$ .
- We choose the critical value such that the **largest** value of the type I error under  $H_0$  is  $\alpha$ .

- That is, we consider the distribution of the test statistic at the boundary of the null hypothesis, i.e.,  $\beta_j^0$ .
- If  $\beta_j = \beta_j^0$ , then the type I error is equal to  $\alpha$ .
- If  $\beta_j < \beta_j^0$ , then the type I error is actually smaller than  $\alpha$ .
- To see this, assume  $\beta_j \leq \beta_j^0$ . Then

$$\frac{\beta_j - \beta_j^0}{\widehat{\sigma}_{\widehat{\beta}_j}} \le 0,$$

and the *t*-statistic

$$t_{\widehat{\beta}_j} = \frac{\widehat{\beta}_j - \beta_j^0}{\widehat{\sigma}_{\widehat{\beta}_j}} = \frac{\widehat{\beta}_j - \beta_j}{\widehat{\sigma}_{\widehat{\beta}_j}} + \frac{\beta_j - \beta_j^0}{\widehat{\sigma}_{\widehat{\beta}_j}} \le \frac{\widehat{\beta}_j - \beta_j}{\widehat{\sigma}_{\widehat{\beta}_j}}$$

• Hence

$$\begin{split} \Pr\left(t_{\widehat{\beta}_{j}} > t_{1-\alpha}^{n-k-1} | H_{0}\right) &= & \Pr\left(\frac{\widehat{\beta}_{j} - \beta_{j}^{0}}{\widehat{\sigma}_{\widehat{\beta}_{j}}} > t_{1-\alpha}^{n-k-1} | H_{0}\right) \\ &= & \Pr\left(\frac{\widehat{\beta}_{j} - \beta_{j}}{\widehat{\sigma}_{\widehat{\beta}_{j}}} + \frac{\beta_{j} - \beta_{j}^{0}}{\widehat{\sigma}_{\widehat{\beta}_{j}}} > t_{1-\alpha}^{n-k-1} | H_{0}\right) \\ &\leq & \Pr\left(\frac{\widehat{\beta}_{j} - \beta_{j}}{\widehat{\sigma}_{\widehat{\beta}_{j}}} > t_{1-\alpha}^{n-k-1} | H_{0}\right) \\ &= & \alpha. \end{split}$$

- That is,  $\alpha$  measures the *minimum degree of protection* against a type I error provided by the test.
- This guarantees strong evidence against  $H_0$  in case of rejection.

- In summary, the procedure of the test is basically the same as before:
- With  $H_0$  being rejected if the *t*-statistic

$$t_{\widehat{\beta}_j} > c = t_{1-\alpha}^{n-k-1},\tag{46}$$

where c is the  $(1-\alpha)-{\rm quantile}$  of the t distribution with n-k-1 degrees of freedom, so that

$$\Pr(t_{\widehat{\beta}_j} > c | \mathsf{H}_0 \text{ is true}) = \alpha.$$

- Note that the critical value is given by the  $(1 \alpha)$ -quantile rather then  $(1 \alpha/2)$ -quantile for one-sided tests.
- The situation for  $H_1: \beta_j > 0$  (and that for  $H_1: \beta_j < 0$ ) is illustrated in the following two figures for  $\alpha = 0.05$  and n k 1 = 18, so that, from Table (2), c = 1.734.



5% rejection rule for the alternative  $H_1$ :  $\beta_i$ >0 with 18 degrees of freedom

• If the alternative hypothesis is, instead of (44)

$$H_1: \beta_j < \beta_j^0, \tag{47}$$

then the null is

$$H_0: \beta_j \ge \beta_j^0, \tag{48}$$

and  $H_0$  is rejected if

$$t_{\widehat{\beta}_j} < c = -t_{1-\alpha}^{n-k-1} = t_{\alpha}^{n-k-1},$$
(49)

i.e., critical value c is the  $\alpha$ -quantile of the t distribution with n-k-1 degrees of freedom, or minus the  $(1-\alpha)$ -quantile of that distribution.



- For example, in the wage equation (41), it may be reasonable to hypothesize that the coefficient of *exper* (i.e.,  $\beta_2$ ) is positive.
- For  $\alpha = 0.05$  and  $\alpha = 0.01$ , the critical values are given by the 0.95 and 0.99 quantiles of the normal distribution, i.e., by 1.645 and 2.326, respectively.
- We have already calculated the  $t\ {\rm statistic}\ {\rm as}$

$$t_{\widehat{\beta}_2} = \frac{0.0041}{0.0017} = 2.412,$$

i.e., we reject  $H_0$  even at the 1% level.

• The p-value for the one-sided alternative can likewise be calculated as

$$p$$
-value =  $\Pr(t_{\hat{\beta}_2} > 2.412 | H_0 \text{ is true}) = 0.0079,$ 

which is just the p-value for the two-sided hypothesis divided by 2.