1

MULTIPLE LINEAR REGRESSION MODEL II: Statistical Inference, Hypothesis Tests, Confidence Intervals

Hüseyin Taştan[1]

[1]Yıldız Technical University
Department of Economics

Econometrics I

2

Multiple Regression Analysis: Inference

- In this class we will learn how to carry out hypothesis tests on population parameters.
- Under the assumption that "Population error term (u) is normally distributed" (MLR.6) we will examine the sampling distributions of OLS estimators.
- First we will learn how to carry out hypothesis tests on single population parameters.
- Then, we will develop testing methods for multiple linear restrictions.
- We will also learn how to decide whether a group of explanatory variables can be excluded from the model.

3

Sampling Distributions of OLS Estimators

- To make statistical inference (hypothesis tests, confidence intervals), in addition to expected values and variances we need to know the sampling distributions of $\hat{\beta}_j$s.
- To do this we need to assume that the error term is normally distributed. Under the Gauss-Markov assumptions the sampling distributions of OLS estimators can have any shape.

Assumption MLR.6 Normality

Population error term u is independent of the explanatory variables and follows a normal distribution with mean 0 and variance σ^2:

$$u \sim N(0, \sigma^2)$$

- Normality assumption is stronger than the previous assumptions.
- Assumption MLR.6 implies that MLR.4, Zero conditional mean, and MLR.5, homoscedasticity, are also satisfied.

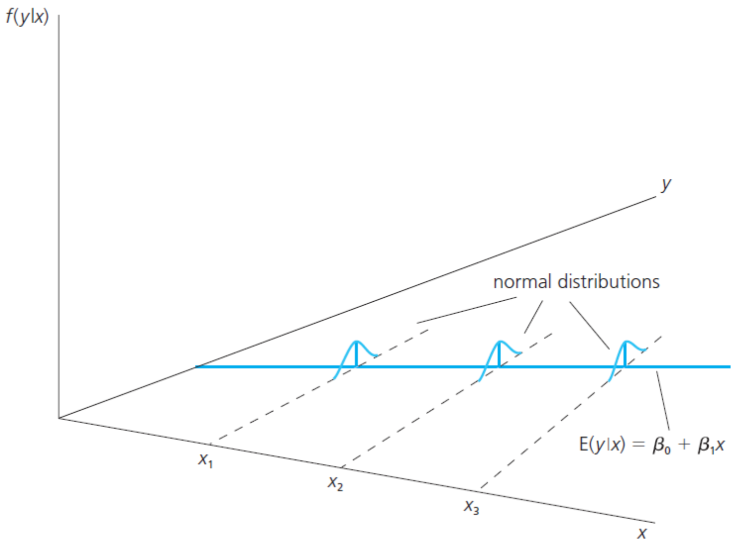
4

Sampling Distributions of OLS Estimators

- Assumptions MLR.1 through MLR.6 are called **classical assumptions**. (Gauss-Markov assumptions + Normality)
- Under the classical assumptions, OLS estimators $\hat{\beta}_j$s are the best unbiased estimators in not only all linear estimators but all estimators (including nonlinear estimators).
- Classical assumptions can be summarized as follows:

$$y|x \sim N(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k, \sigma^2)$$

Normality assumptions in the simple regression



How can we justify the normality assumption?

- u is the sum of many different unobserved factors affecting y.
- Therefore, we can invoke the Central Limit Theorem (CLT) to conclude that u has an approximate normal distribution.
- CLT assumes that unobserved factors in u affect y in an additive fashion.
- If u is a complicated function of unobserved factors then the CLT may not apply.
- Normality: usually an empirical matter.
- In some cases, normality assumption may be violated, for example, distribution of wages may not be normal (positive values, minimum wage laws, etc.). In practice, we assume that conditional distribution is close to being normal.
- In some cases, transformations of variables (e.g., natural log) may yield an approximately normal distribution.

Sampling Distributions of OLS Estimators

Normal Sampling Distributions

Under the assumptions MLR.1 through MLR.6 OLS estimators follow a normal distributions (conditional on xs):

$$\hat{\beta}_j \sim N\left(\beta_j, \text{Var}(\hat{\beta}_j)\right)$$

Standardizing we obtain:

$$\frac{\hat{\beta}_j - \beta_j}{sd(\hat{\beta}_j)} \sim N(0,1)$$

OLS estimators can be written as a linear combination of error terms. Recall that linear combinations of normally distributed random variables also follow normal distribution.

Testing Hypotheses about a Single Population Parameter: The t Test

$$\frac{\hat{\beta}_j - \beta_j}{sd(\hat{\beta}_j)} \sim N(0,1)$$

- Replacing the standard deviation (sd) in the denominator by standard error (se):

$$\frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \sim t_{n-k-1}$$

- The t test is used in testing hypotheses about a single population parameter as in $H_0 : \beta_j = \beta_j^*$.

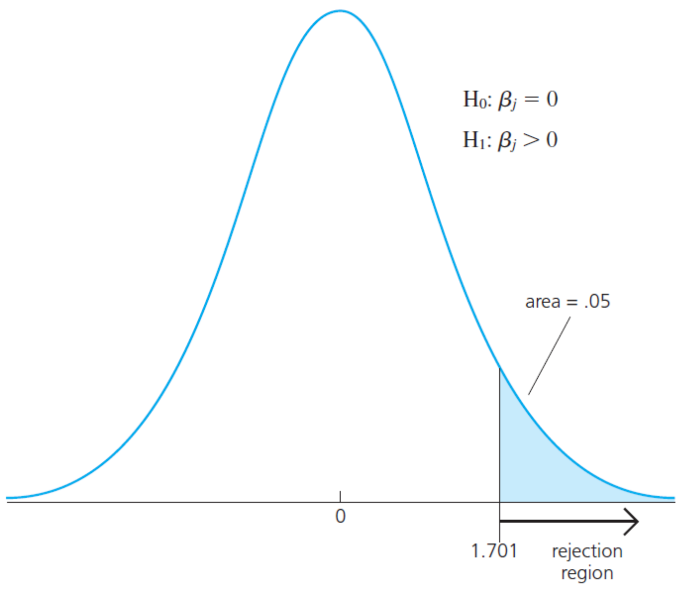
The t Test

Testing Against One-Sided Alternatives (Right Tail)

$$H_0 : \beta_j = 0$$
$$H_1 : \beta_j > 0$$

- The meaning of the null hypothesis: after controlling for the impacts of $x_1, x_2, \ldots, x_{j-1}, x_{j+1}, \ldots, x_k$, x_j has **no effect** on the expected value of y.
- Test statistic
$$t_{\hat{\beta}_j} = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \sim t_{n-k-1}$$
- Decision Rule: reject the null hypothesis if $t_{\hat{\beta}_j}$ is larger than the $100\alpha\%$ critical value associated with t_{n-k-1} distribution.
$$\text{If } \ t_{\hat{\beta}_j} > c, \quad \text{then } \ \text{REJECT } \ H_0$$
otherwise fail to reject H_0

5% Decision Rule with dof=28



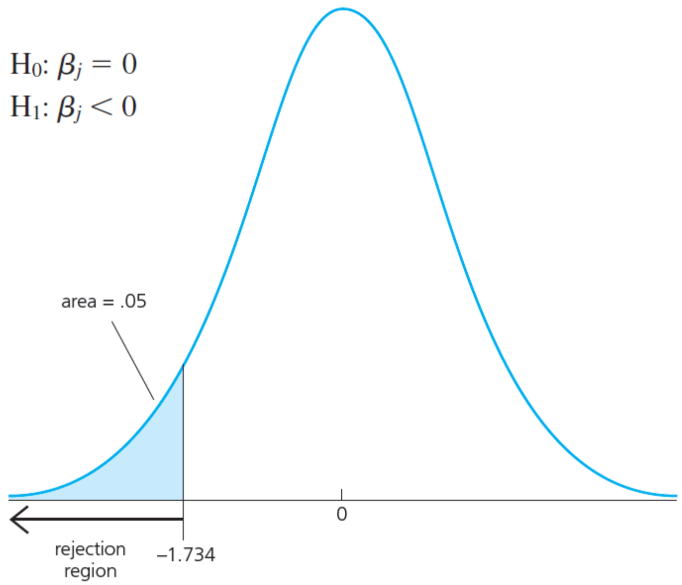
The t Test

Testing Against One-Sided Alternatives (Left Tail)

$$H_0 : \beta_j = 0$$
$$H_1 : \beta_j < 0$$

- The test statistic:
$$t_{\hat{\beta}_j} = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \sim t_{n-k-1}$$
- Decision Rule: If calculated test statistic $t_{\hat{\beta}_j}$ is smaller than the critical value at the chosen significance level we reject the null hypothesis:
$$\text{If } \ t_{\hat{\beta}_j} < -c, \quad \text{then } \ \text{REJECT } \ H_0$$
otherwise, fail to reject H_0.

Decision Rule for the left tail test, dof=18



The t Test

Testing Against Two-Sided Alternatives

$$H_0 : \beta_j = 0$$
$$H_1 : \beta_j \neq 0$$

- The test statistic:

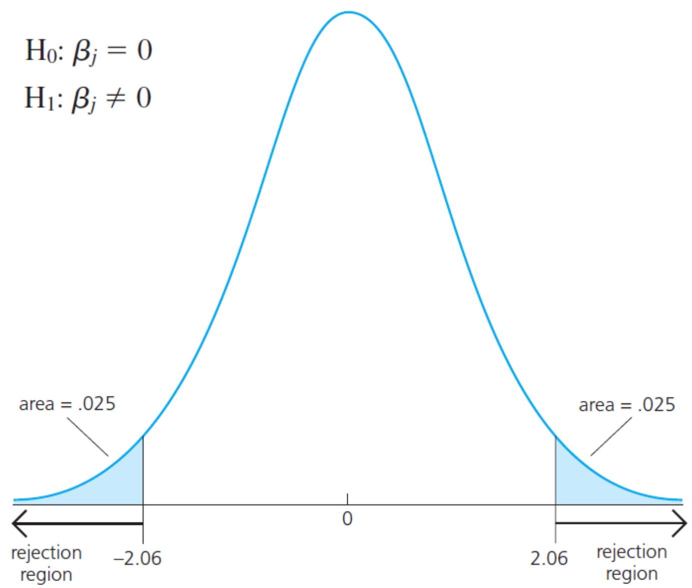
$$t_{\hat{\beta}_j} = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \sim t_{n-k-1}$$

- Decision Rule: If the absolute value of the test statistic $|t_{\hat{\beta}_j}|$ is larger than the critical value at the $100\alpha/2$ significance level ($c = t_{n-k-1,\alpha/2}$) then we reject H_0:.

$$\text{If} \quad |t_{\hat{\beta}_j}| > c, \quad \text{then} \quad \text{REJECT} \quad H_0$$

otherwise, we fail to reject.

Decision Rule for Two-Sided Alternatives at 5% Significance Level, dof=25



R Example: Wage equation

```
 > res1 <- lm(log(wage) ~ educ + exper + tenure, data = wage1)
> summary(res1)

Call:
lm(formula = log(wage) ~ educ + exper + tenure, data = wage1)

Residuals:
     Min       1Q   Median       3Q      Max
-2.05802 -0.29645 -0.03265  0.28788  1.42809

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.284360   0.104190   2.729  0.00656 **
educ        0.092029   0.007330  12.555  < 2e-16 ***
exper       0.004121   0.001723   2.391  0.01714 *
tenure      0.022067   0.003094   7.133 3.29e-12 ***
---
Residual standard error: 0.4409 on 522 degrees of freedom
Multiple R-squared:  0.316,Adjusted R-squared:  0.3121
F-statistic: 80.39 on 3 and 522 DF,  p-value: < 2.2e-16
```

The t Test: Examples

Log-level Wage Equation: wage1.gdt

$$\widehat{\log(\text{wage})} = \underset{(0.104)}{0.284} + \underset{(0.007)}{0.092}\,\text{educ} + \underset{(0.0017)}{0.004}\,\text{exper} + \underset{(0.003)}{0.022}\,\text{tenure}$$
$$n = 526 \quad R^2 = 0.316$$

(standard errors in parentheses)

- Is exper statistically significant? Test $H_0 : \beta_{exper} = 0$ against $H_1 : \beta_{exper} > 0$
- The t-statistic is: $t_{\hat{\beta}_j} = 0.004/0.0017 = 2.41$
- One-sided critical value at 5% significance level is $c_{0.05} = 1.645$, at 1% level $c_{0.01} = 2.326$, dof = 526-4=522
- Since $t_{\hat{\beta}_j} > 2.326$ we reject H_0. Exper is statistically significant at 1% level.
- $\hat{\beta}_{exper}$ is statistically greater than zero at the 1% significance level.

The t Test: Examples

Student Performance and School Size: meap93.gdt

$$\widehat{\text{math10}} = \underset{(6.114)}{2.274} + \underset{(0.0001)}{0.00046}\,\text{totcomp} + \underset{(0.0398)}{0.048}\ \text{staff} - \underset{(0.00022)}{0.0002}\ \text{enroll}$$
$$n = 408 \quad R^2 = 0.0541$$

math10: mathematics test results (a measure of student performance), totcomp: total compensation for teachers (a measure of teacher quality), staff: number of staff per 1000 students (a measure of how much attention students get), enroll: number of students (a measure of school size)

- Test $H_0 : \beta_{enroll} = 0$ against $H_1 : \beta_{enroll} < 0$
- Calculated t-statistic: $t_{\hat{\beta}_j} = -0.0002/0.00022 = -0.91$
- One-sided critical value at the 5% significance level: $c_{0.05} = -1.645$
- Since $t_{\hat{\beta}_j} > -1.645$ we fail to reject H_0.
- $\hat{\beta}_{enroll}$ is statistically insignificant (not different from zero) at the 5% level.

The t Test: Examples

Student Performance and School Size: Level-Log model

$$\widehat{\text{math10}} = -\underset{(48.7)}{207.67} + \underset{(4.06)}{21.16}\,\log(\text{totcomp}) + \underset{(4.19)}{3.98}\ \log(\text{staff}) - \underset{(0.69)}{1.27}\ \log(\text{enroll})$$
$$n = 408 \quad R^2 = 0.065$$

- Test $H_0 : \beta_{log(enroll)} = 0$ against $H_1 : \beta_{log(enroll)} < 0$
- Calculated t-statistic: $t_{\hat{\beta}_j} = -1.27/0.69 = -1.84$
- Critical value at 5%: $c_{0.05} = -1.645$
- Since $t_{\hat{\beta}_j} < -1.645$ we reject H_0 in favor of H_1.
- $\hat{\beta}_{log(enroll)}$ is statistically significant at the 5% significance level (smaller than zero).

The t Test: Examples

Determinants of College GPA, gpa1.gdt

$$\widehat{\text{colGPA}} = \underset{(0.331)}{1.389} + \underset{(0.094)}{0.412}\,\text{hsGPA} + \underset{(0.011)}{0.015}\ \text{ACT} - \underset{(0.026)}{0.083}\,\text{skipped}$$
$$n = 141 \quad R^2 = 0.23$$

skipped: average number of lectures missed per week

- Which variables are statistically significant using two-sided alternative?
- Two-sided critical value at the 5% significance level is $c_{0.025} = 1.96$. Because dof=141-4=137 we can use standard normal critical values.
- $t_{hsGPA} = 4.38$: hsGPA is statistically significant.
- $t_{ACT} = 1.36$: ACT is statistically insignificant.
- $t_{skipped} = -3.19$: skipped is statistically significant at the 1% level ($c = 2.58$).

Testing Other Hypotheses about β_j

The t Test

Null hypothesis is

$$H_0 : \beta_j = a_j$$

test statistic is

$$t = \frac{\hat{\beta}_j - a_j}{se(\hat{\beta}_j)} \sim t_{n-k-1}$$

or

$$t = \frac{estimate - hypothesized\ value}{standard\ error}$$

- t statistic measures how many estimated standard deviations $\hat{\beta}_j$ is away from the hypothesized value.
- Depending on the alternative hypothesis (left tail, right tail, two-sided) the decision rule is the same as before.

Testing Other Hypotheses about β_j: Example

Campus crime and university size: campus.gdt

$$crime = \exp(\beta_0) enroll^{\beta_1} \exp(u)$$

Taking natural log:

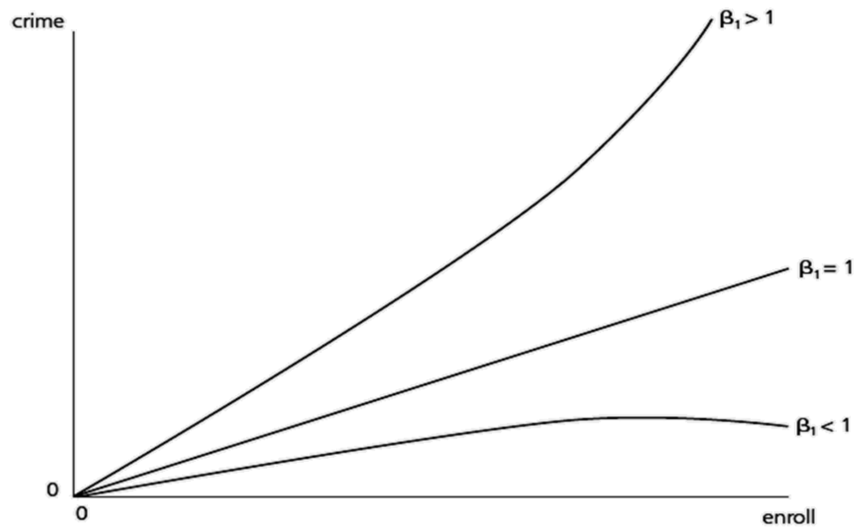
$$\log(crime) = \beta_0 + \beta_1 \log(enroll) + u$$

- Data set: contains annual number of crimes and enrollment for 97 universities in USA
- We wan to test:

$$H_0 : \beta_1 = 1$$

$$H_1 : \beta_1 > 1$$

Crime and enrollment: $crime = enroll^{\beta_1}$



Testing Other Hypotheses about β_j: Example

Campus crime and enrollment: campus.gdt

$$\widehat{\log(\text{crime})} = \underset{(1.03)}{-6.63} + \underset{(0.11)}{1.27} \log(\text{enroll})$$

$$n = 97 \quad R^2 = 0.585$$

- Test: $H_0 : \beta_1 = 1$, $H_1 : \beta_1 > 1$
- Calculated test statistic

$$t = \frac{1.27 - 1}{0.11} \approx 2.45 \sim \ t_{95}$$

- Critical value at the 5% significance level: c=1.66 ($dof = 120$), thus we reject H_0.
- Can we say that we measured the ceteris paribus effect of university size? What other factors should we consider?

Testing Other Hypotheses about β_j: Example

Housing Prices and Air Pollution: hprice2.gdt

Dependent variable: log of the median house price (log(price))
Explanatory variables:
log(nox): the amount of nitrogen oxide in the air in the community,
log(dist): distance to employment centers,
rooms: average number of rooms in houses in the community,
stratio: average student-teacher ratio of schools in the community

- Test: $H_0 : \beta_{log(nox)} = -1$ against $H_1 : \beta_{log(nox)} \neq -1$
- Estimated value: $\hat{\beta}_{log(nox)} = -0.954$, standard error $= 0.117$
- Test statistic:

$$t = \frac{-0.954 - (-1)}{0.117} = \frac{-0.954 + 1}{0.117} \approx 0.39 \sim \ t_{501} \sim \ N(0, 1)$$

- Two-sided critical value at the 5% significance level is c=1.96. Thus, we fail to reject H_0.

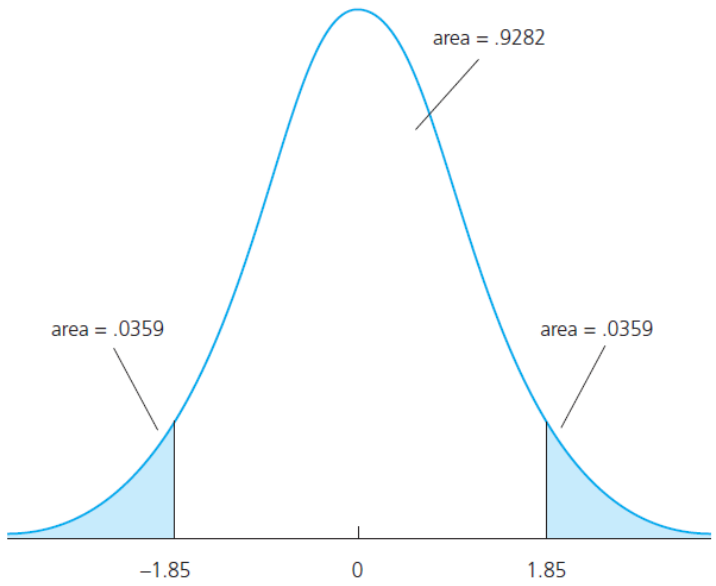
Computing p-values for t-tests

- Instead of choosing a significance level (e.g. 1%, 5%, 10%), we can compute the smallest significance level at which the null hypothesis would be rejected.
- This is called p-value.
- In standard regression softwares p-values are reported for $H0 : \beta_j = 0$ against two-sided alternative.
- In this case, p-value gives us the probability of drawing a number from the t distribution which is larger than the absolute value of the calculated t-statistic:

$$P(|T| > |t|)$$

- The smaller the p-value the greater the evidence against the null hypothesis.

p-value: Graphical depiction



Large Standard Errors and Small t Statistics

- As the sample size (n) gets bigger the standard errors of $\hat{\beta}_j$s become smaller.
- Therefore, as n becomes larger it is more appropriate to use small significance levels (such as 1%).
- One reason for large standard errors in practice may be due to high collinearity among explanatory variables (multicollinearity).
- If explanatory variables are highly correlated it may be difficult to determine the partial effects of variables.
- In this case the best we can do is to collect more data.

Guidelines for Economic and Statistical Significance

- Check for statistical significance: if significant discuss the practical and economic significance using the magnitude of the coefficient.
- If a variable is not statistically significant at the usual levels (1%, 5%, 10%) you may still discuss the economic significance and statistical significance using p-values.
- Small t-statistics and wrong signs on coefficients: these can be ignored in practice, they are statistically insignificant.
- A significant variable that has the unexpected sign and practically large effect is much more difficult to interpret. This may imply a problem associated with model specification and/or data problems.

29

Confidence Intervals

- We know that:

$$t_{\hat{\beta}_j} = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \sim t_{n-k-1}$$

- Using this ratio we can construct the $\%100(1-\alpha)$ confidence interval:

$$\hat{\beta}_j \pm c \cdot \mathrm{se}(\hat{\beta}_j)$$

- Lower and upper bounds of the confidence interval are:

$$\underline{\beta_j} \equiv \hat{\beta}_j - c \cdot \mathrm{se}(\hat{\beta}_j), \quad \overline{\beta_j} \equiv \hat{\beta}_j + c \cdot \mathrm{se}(\hat{\beta}_j)$$

30

Confidence Intervals

$$[\hat{\beta}_j - c \cdot \mathrm{se}(\hat{\beta}_j),\ \hat{\beta}_j + c \cdot \mathrm{se}(\hat{\beta}_j)]$$

- How do we interpret confidence intervals?
- If random samples were obtained over and over again and confidence intervals are computed for each sample then the unknown population value β_j would lie in the confidence interval for $100(1-\alpha)\%$ of the samples.
- For example we would say 95 of the confidence intervals out of 100 would contain the true value. Note that $\alpha/2 = 0.025$ in this case.
- In practice, we only have one sample and thus only one confidence interval estimate. We do not know if the estimated confidence interval contains the true value.

31

Confidence Intervals

- We need three quantities to calculate confidence intervals. coefficient estimate, standard error and critical value.
- For example, for dof=25 and 95% confidence level, confidence interval for a population parameter can be calculated using:

$$[\hat{\beta}_j - 2.06 \cdot \mathrm{se}(\hat{\beta}_j),\ \hat{\beta}_j + 2.06 \cdot \mathrm{se}(\hat{\beta}_j)]$$

- If $n-k-1 > 50$ then 95% confidence interval can easily be calculated using $\hat{\beta}_j \pm 2 \cdot \mathrm{se}(\hat{\beta}_j)$.
- Suppose we want to test the following hypothesis:

$$H_0 : \beta_j = a_j$$

$$H_1 : \beta_j \neq a_j$$

- We reject H_0 at the 5% significance level in favor of H_1 iff the 95% confidence interval does **not** contain a_j.

32

Example: Hedonic Price Model for Houses

- A hedonic price model relates the price to the product's characteristics.
- For example, in a hedonic price model for computers the price of computers is regressed on the physical characteristics such as CPU power, RAM size, notebook/desktop, etc.
- Similarly, the value of a house is determined by several characteristics: size, number of rooms, distance to employment centers, schools and parks, crime rate in the community, etc.
- Dependent variable: log(price)
- Explanatory variables: sqrft (square footage, size) (1 square foot = 0.09290304 m^2, 100m^2 $\approx$ 1076 ftsq; bdrms: number of rooms, bthrms: number of bathrooms.

Example: Hedonic Price Model for Houses

Estimation Results

$$\widehat{\log(\text{price})} = \underset{(1.15)}{7.46} + \underset{(0.184)}{0.634}\,\mathsf{log(sqrft)} - \underset{(0.059)}{0.066}\,\mathsf{bdrms} + \underset{(0.075)}{0.158}\,\mathsf{bthrms}$$

$$n = 19 \quad R^2 = 0.806$$

- Both price and sqrft are in logs, therefore the coefficient estimate gives us elasticity: Holding bdrms and bthrms fixed, if the size of the house increases 1% then the value of the house is predicted to increase by 0.634%.
- dof=n-k-1=19-3-1=15 critical value for t_{15} distribution c=2.131 using $\alpha = 0.05$. Thus, 95% confidence interval is

$$0.634 \pm 2.131 \cdot (0.184) \Rightarrow [0.242, 1.026]$$

- Because this interval does not contain zero, we reject the null hypothesis that the population parameter is insignificant.
- The coefficient estimate on the number of rooms is $(-)$. Why?

Example: Hedonic Price Model for Houses

Estimation Results

$$\widehat{\log(\text{price})} = \underset{(1.15)}{7.46} + \underset{(0.184)}{0.634}\,\mathsf{log(sqrft)} - \underset{(0.059)}{0.066}\,\mathsf{bdrms} + \underset{(0.075)}{0.158}\,\mathsf{bthrms}$$

$$n = 19 \quad R^2 = 0.806$$

- 95% confidence interval for β_{bdrms} is $[-0.192,\ 0.006]$.
- This interval contains 0. Thus, its effect is statistically insignificant.
- Interpretation of coefficient estimate on bthrms: ceteris paribus, if the number of bathrooms increases by 1, house prices are predicted to increase by approximately 100(0.158)%=15.8% on average.
- 95% confidence interval is $[-0.002, 0.318]$. Technically this interval does not contain zero but the lower confidence limit is close to zero. It is better to compute p-value.