4. Statistical Inference

4.1 Sampling Distributions of the OLS Estimators

Regression model

(1) $$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + u_i.$$

To the assumptions 1–5 we add

Assumption 6: The error component u is independent of $x_1, \ldots x_k$ and

(2) $$u \sim N(0, \sigma_u^2).$$

Remark 4.1: Assumption 6 implies
$\mathbb{E}[u|x_1, \ldots .x_k] = \mathbb{E}[u] = 0$ (Assumption 4) and
$\mathbb{V}\text{ar}[u|x_1, \ldots, x_k] = \mathbb{V}\text{ar}[u] = \sigma_u^2$ (Assumption 5).

Remark 4.2: Assumption 2, i.e., $\mathbb{C}\text{ov}[u_i, u_j] = 0$ together with Assumption 6 implies that $u_1, \ldots, u_n$ are independent.

Remark 4.3: Under assumptions 1–6 the OLS estimators $\widehat{\beta}_1, \ldots, \widehat{\beta}_k$ are Minimum Variance Unbiased Estimators (MVUE). That is they are best among all unbiased estimators (not only linear).

Remark 4.4:

$$y|\mathbf{x} \sim N(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k, \sigma_u^2), \tag{1}$$

where $\mathbf{x} = (x_1, \ldots, x_k)'$ and $y|\mathbf{x}$ means "conditional on $\mathbf{x}$".

Theorem 4.1: Under the assumptions 1–6, conditional on the sample values of the explanatory variables

$$\hat{\beta}_j \sim N(\beta_j, \sigma^2_{\hat{\beta}_j}) \tag{4}$$

and therefore

$$\frac{\hat{\beta}_j - \beta_j}{\sigma_{\hat{\beta}_j}} \sim N(0, 1), \tag{5}$$

where $\sigma^2_{\hat{\beta}_j} = \mathrm{Var}[\hat{\beta}_j]$ and $\sigma_{\hat{\beta}_j} = \sqrt{\mathrm{Var}[\hat{\beta}_j]}$.

4.2 Testing for single population coefficients, the t-test

Theorem 4.2: Under the assumptions 1–6

$$\frac{\hat{\beta}_j - \beta_j}{s_{\hat{\beta}_j}} \sim t_{n-k-1}, \tag{6}$$

(the t-distribution with $n-k-1$ degrees of freedom) where $s_{\hat{\beta}_j} = \text{se}(\hat{\beta}_j)$ and $k+1$ is the number of estimated regression coefficients.

Remark 4.5: The only difference between (5) and (6) is that in the latter the standard deviation parameter $\sigma_{\hat{\beta}_j}$ is replaced by its estimator $s_{\hat{\beta}_j}$.

In most applications the interest lies in testing the *null hypothesis*:

(7) $$H_0 : \beta_j = 0.$$

The t-test statistic is

(8) $$t_{\hat{\beta}_j} = \frac{\hat{\beta}_j}{s_{\hat{\beta}_j}},$$

which is t-distributed with $n - k - 1$ degrees of freedom *if the null hypothesis is true*.

These "t-ratios" are printed in standard computer output in regression applications.

Example 4.1: Wage example computer output.

Dependent Variable: LOG(WAGE)
Method: Least Squares
Sample: 1 526
Included observations: 526

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.583773	0.097336	5.997510	0.0000
EDUC	0.082744	0.007567	10.93534	0.0000

R-squared	0.185806	Mean dependent var	1.623268
Adjusted R-squared	0.184253	S.D. dependent var	0.531538
S.E. of regression	0.480079	Akaike info criterion	1.374061
Sum squared resid	120.7691	Schwarz criterion	1.390279
Log likelihood	-359.3781	F-statistic	119.5816
Durbin-Watson stat	1.801328	Prob(F-statistic)	0.000000

Remark 4.6: Hypothesis tests are always about population parameters. It never makes sense to state null hypothesis like "$H_0 : \widehat{\beta} = 0$"!

Testing Against One-Sided Alternatives

One-sided alternatives

$$H_1 : \beta_j > 0 \tag{9}$$

or

$$H_1 : \beta_j < 0. \tag{10}$$

In the former the rejection rule is to reject the null hypothesis at the chosen significance level, α

$$t_{\hat{\beta}_j} > c_\alpha, \tag{11}$$

where c_α is the $1-\alpha$ fractile (or percentile) from the t-distribution with $n-k-1$ degrees of freedom, such that $P(t_{\hat{\beta}_j} > c_\alpha | H_0 \text{ is true}) = \alpha$. α is called the significance level of the test. Typically α is 0.05 or 0.01, i.e., 5% or 1%.

In the case of (10) the H_0 is rejected if

$$t_{\hat{\beta}_j} < -c_\alpha. \tag{12}$$

These tests are one-tailed test.

Example 4.2: In the wage example, test

$$H_0 : \beta_{\text{exper}} = 0$$

against

$$H_1 : \beta_{\text{exper}} > 0.$$

$\widehat{\beta}_{\text{exper}} = 0.004121$, $s_{\widehat{\beta}_{\text{exper}}} = 0.001723$. Thus

$$t_{\widehat{\beta}_{\text{exper}}} = \frac{\widehat{\beta}_{\text{exper}}}{s_{\widehat{\beta}_{\text{exper}}}} = \frac{0.004121}{0.001723} \approx 2.391.$$

Looking up in a table, we would find that $c_{0.01} \approx 2.33$ and $c_{0.005} \approx 2.58$. We may thus reject $H_0 : \beta_{\text{exper}} = 0$ against $H_1 : \beta_{\text{exper}} > 0$ at a significance level of 1% but not at 0.5%, since $2.33 < 2.39 < 2.58$.

Two-Sided Alternatives

If the null hypothesis is $H_0 : \beta_j = 0$, the two-sided alternative is

$$H_1 : \beta_j \neq 0. \tag{13}$$

The null hypothesis is rejected at the significance level α if

$$|t_{\hat{\beta}_j}| > c_{\alpha/2}. \tag{14}$$

Example 4.2: (continued) Looking up in a table, we find $c_{0.02/2} = 2.326$ and $c_{0.01/2} = 2.576$. We may thus reject $H_0 : \beta_{\text{exper}} = 0$ against $H_1 : \beta_{\text{exper}} \neq 0$ at least at a significance level of 2% (but not, as in the one-sided test, at $\alpha = 1\%$).

Other Hypotheses About β_j

Generally the null hypothesis can be also

$$H_0 : \beta_j = \beta_j^*, \tag{15}$$

where β_j^* is some given value (for example $\beta_j^* = 1$, so $H_0 : \beta_j = 1$).

The test statistic is again a t-statistic

$$t = \frac{\hat{\beta}_j - \beta_j^*}{s_{\hat{\beta}_j}}. \tag{16}$$

Under the null hypothesis (15) the test statistic (16) is again t-distributed with $n - k - 1$ degrees of freedom.

Remark 4.7: The computer print outs always give the t-ratios, i.e., test against zero. Consequently, they cannot be used to test the more general hypothesis (15). You can, however, use the standard errors and compute the test statistics of the form (16).

Example 4.2 (continued): Test

$$H_0 : \beta_{\text{exper}} = 0.005$$

against

$$H_1 : \beta_{\text{exper}} \neq 0.005.$$

$\widehat{\beta}_{\text{exper}} = 0.004121$, $s_{\widehat{\beta}_{\text{exper}}} = 0.001723$. Thus

$$t = \frac{\widehat{\beta}_{\text{exper}} - \beta^*_{\text{exper}}}{s_{\widehat{\beta}_{\text{exper}}}} = \frac{0.004121 - 0.005}{0.001723} \approx -0.51.$$

Looking up in a table, we would find that $c_{0.2/2} \approx 1.28 > |-0.51|$. We are thus unable to reject $H_0 : \beta_{\text{exper}} = 0.005$ against $H_1 : \beta_{\text{exper}} \neq 0.005$ even at a significance level of 20%. So there is no evidence against the hypothesis, that an extra year working experience yields 0.5% more salary (everything else being equal).

Example 4.3: Housing prices and air pollution.

A sample of 506 communities in Boston area.
Variables:
price (y) = median housing price
nox (x_1) = nitrogen oxide, parts per 100 mill.
dist (x_2) = weighted dist. to 5 employ centers
rooms (x_3) = avg number of rooms per house
stratio (x_4) = average student-teacher ratio of schools in community

Specified model

(17) $\log(y) = \beta_0 + \beta_1 \log(x_1) + \beta_2 \log(x_2) + \beta_3 x_3 + \beta_4 x_4 + u$

β_1 is the price elasticity of nox. We wish to test

$$H_0 : \beta_1 = -1$$

against

$$H_1 : \beta_1 \neq -1.$$

Estimation results:

Dependent Variable: LOG(PRICE)
Method: Least Squares
Sample: 1 506
Included observations: 506

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	11.08386	0.318111	34.84271	0.0000
LOG(NOX)	-0.953539	0.116742	-8.167932	0.0000
LOG(DIST)	-0.134339	0.043103	-3.116693	0.0019
ROOMS	0.254527	0.018530	13.73570	0.0000
STRATIO	-0.052451	0.005897	-8.894399	0.0000

R-squared	0.584032	Mean dependent var	9.941057
Adjusted R-squared	0.580711	S.D. dependent var	0.409255
S.E. of regression	0.265003	Akaike info criterion	0.191679
Sum squared resid	35.18346	Schwarz criterion	0.233444
Log likelihood	-43.49487	F-statistic	175.8552
Durbin-Watson stat	0.681595	Prob(F-statistic)	0.000000

$$t = \frac{-0.953539 - (-1)}{0.116742} = \frac{-0.953539 + 1}{0.116742} \approx 0.393.$$

$t_{501}(0.025) \approx z(0.025) = 1.96$, which is far higher than the test statistic. Thus we do not reject the null hypothesis and conclude that there is not empirical evidence that the elasticity would differ from -1.

p-values

The p-value is defined as the smallest significance level at which the null-hypothesis could be rejected.

Thus we can base our inference on the p-value instead of finding from the tables the critical values. The decision rule simply is that if the p-value is smaller than the selected significance level α we reject the null hypothesis.

Technically the p-value is calculated as the probability

$$p = \begin{cases} P(T > t_{\text{obs}}|H_0), & \text{if the alternative hypothesis is } H_1: \beta > \beta^* \\ P(T < t_{\text{obs}}|H_0), & \text{if the alternative hypothesis is } H_1: \beta < \beta^* \\ P(|T| > t_{\text{obs}}|H_0), & \text{if the alternative hypothesis is } H_1: \beta \neq \beta^* \end{cases} \tag{18}$$

where T is a t-distributed random variable and t_{obs} is the value of t-statistic calculated form the sample (observed t-statistic).

Remark 4.8: The computer output contains p-values for the null hypothesis that the coefficient is zero and the alternative hypothesis is that it differs form zero (two-sided).

Example 4.4: In the previous example the p-values indicate that all the coefficient estimates differ (highly) statistically significantly from zero.

For the null hypothesis $H_0 : \beta_1 = -1$ with the alternative hypothesis $H_1 : \beta_1 \neq -1$ p-value is obtained by using the standardized normal distribution as

$$2(1 - \Phi(0.398)) \approx 0.69,$$

where $\Phi(z)$ is the cumulative distribution function of the standardized normal distribution.

4.3 Confidence Intervals for the Coefficients

From the fact that

$$(19) \qquad \frac{\widehat{\beta}_j - \beta_j}{s_{\widehat{\beta}_j}} \sim t_{n-k-1}$$

we get for example a 95% confidence interval for the unknown parameter β_j as

$$(20) \qquad \widehat{\beta}_j \pm c_{\frac{1}{2}\alpha} s_{\widehat{\beta}_j},$$

where $c_{\alpha/2}$ is again the $1-\alpha/2$ fractile of the appropriate t-distribution.

Interpretation!

Example 4.5: The 95% confidence interval for $\beta_{\text{nox}} = \beta_1$ is

$$\begin{aligned} \widehat{\beta}_{\text{nox}} \pm c_{.025} s_{\widehat{\beta}_{nox}} &= -0.953539 \pm 1.96 \times 0.116742 \\ &= -0.953539 \pm 0.22881432 \end{aligned} \tag{21}$$

or

$$(-1.182, -0.725). \tag{22}$$

We observe that $-1 \in (-1.182, -0.725)$.

The F-test

Hypotheses $H_0 : \beta_j = 0$ test whether a single coefficient is zero, i.e. whether variable x_j has marginal impact on y.

Hypothesis

(23) $$H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0$$

tests whether none of the x-variables affect y. I.e., whether the model is

$$y = \beta_0 + u$$

instead of

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u.$$

The alternative hypothesis is

(24) $$H_1 : \text{ at least one } \beta_j \neq 0.$$

Null hypothesis (23) is tested by the F-statistic, called the *F-statistic for overall significance of a regression*

(25)

$$F = \frac{\mathrm{SSE}/k}{\mathrm{SSR}/(n-k-1)} = \frac{R^2/k}{(1-R^2)/(n-k-1)},$$

which under the null hypothesis is F-distributed with k and $n-k-1$ degrees of freedom.

This is again printed in the standard computer output of regression analysis.

Example 4.6 In the house price example $F = 175.8552$ with p-value 0.0000, which is highly significant as would be expected.

The principle of the F-test can be used to test more general (linear) hypotheses.

For example to test whether the last q variables contribute y, the null hypothesis is

$$H_0 : \beta_{k-q+1} = \beta_{k-q+2} = \cdots = \beta_k = 0. \tag{26}$$

The restricted model satisfying the null hypothesis is

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_{k-q} x_{k-q} + u \tag{27}$$

with $k - q$ explanatory variables, and the unrestricted model is

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u \tag{28}$$

with k explanatory variables. Thus the restricted model is a special case of the unrestricted one.

The F-statistic is

$$F = \frac{(\mathrm{SSR}_r - \mathrm{SSR}_{ur})/q}{\mathrm{SSR}_{ur}/(n-k-1)}, \tag{29}$$

where SSR_r is the residual sum of squares from the restricted model (27) and SSR_u is the residual sum of squares for the unrestricted model (28).

Under the null hypothesis the test statistic (29) is again F-distributed with $q = df_r - df_{ur}$ and $n - k - 1$ degrees of freedom, where df_r is the degrees of freedom of SSR_r and df_{ur} is the degrees of freedom of SSR_{ur}.

Remark 4.9: Testing for single regression parameters is a special case of (26), and it can be shown that in such a case the F-statistic from (29) equals $t^2_{\hat{\beta}_j}$ with identical p-values for the F- and the t-test.

Remark 4.10: It can be easily shown that

$$F = \frac{(R^2_{ur} - R^2_r)/q}{(1 - R^2_{ur})/(n - k - 1)}, \tag{30}$$

where R^2_{ur} and R^2_r are the R-squares of the unrestricted and restricted models, respectively.

Testing General Linear Restrictions

The principle used in constructing the F-test in (29) can be extended for testing general linear restrictions between the parameters.

As an example, consider the regression model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u. \tag{31}$$

If the hypothesis is

$$H_0 : \beta_1 + \beta_2 + \beta_3 = 1, \tag{32}$$

we can set, for example $\beta_3 = 1 - \beta_1 - \beta_2$, such that in the restricted model under H_0:

(33)

$$\begin{aligned} y =& \beta_0 + \beta_1 x_1 + \beta_2 x_2 + (1 - \beta_1 - \beta_2) x_3 + u \\ =& \beta_0 + \beta_1 (x_1 - x_3) + \beta_2 (x_2 - x_3) + x_3 + u. \end{aligned}$$

In the restricted model, we can estimate β_1 and β_2 from

(34)

$$\underbrace{y - x_3}_{\tilde{y}} = \beta_0 + \beta_1 \underbrace{(x_1 - x_3)}_{\tilde{x}_1} + \beta_2 \underbrace{(x_2 - x_3)}_{\tilde{x}_2} + u$$

and calculate the residual sum of squares for the restricted model,

$$\text{RSS}_r = \sum_{i=1}^{n} (y_i - \tilde{y}_i)^2 \tag{35}$$

from the estimates $\tilde{\beta}_1$ and $\tilde{\beta}_2$, which we then compare by using the F-statistic (29) with the residual sum of squares for the unrestricted model (31).

In the restricted model one parameter less is estimated than in the unrestricted case. Thus the degrees of freedom in the F-statistic are 1 and $n - k - 1$

4.3 On Reporting the Regression Results

(1) Estimated coefficients and interpret them

(2) Standard errors (or if t-ratios or p-values)

(3) R-squared and number of observations

(4) Optionally, standard error of regression