EC508: Econometrics
Measures of Fit and Hypothesis Testing

Jean-Jacques Forneron

Spring, 2023

Boston University

Measures of fit for multiple regression (SW 6.4)

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \cdots + \beta_k X_{k,i} + u_i, \quad i = 1, \ldots, n$$

- Actual = predicted + residual: $Y_i = \hat{Y}_i + \hat{u}_i$
- SER = std. deviation of $\hat{u}_i$ (with d.f. correction)
- RMSE = std. deviation of $\hat{u}_i$ (without d.f. correction)
- R^2 = fraction of variance of Y explained by X
- $\bar{R}^2$ = "adjusted-R^2" = R^2 with a degrees-of-freedom correction that adjusts for estimation uncertainty; $\bar{R}^2 < R^2$

SER and RMSE

- SER and RMSE are measures of the spread of the Ys around the regression line:

$$SER = \sqrt{\frac{1}{n-k-1}\sum_{i=1}^{n}\hat{u}_i^2}, \quad RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\hat{u}_i^2}$$

R^2 and $\bar{R}^2$

- The R^2 is the fraction of the variance explained – same definition as in regression with a single regressor:

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

where

$$ESS = \sum_{i=1}^{n} (\hat{Y}_i - \hat{Y})^2$$

$$SSR = \sum_{i=1}^{n} \hat{u}_i^2$$

$$TSS = \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

- The R^2 always increases when you add another regressor (why?) – a bit of a problem for a measure of “fit”

R^2 and $\bar{R}^2$, cont'd

- The $\bar{R}^2$ (the "adjusted R^2") corrects this problem by "penalizing" you for including another regressor – the $\bar{R}^2$ does not necessarily increase when you add another regressor.
- Adjusted-R^2:
$$\bar{R}^2 = 1 - \frac{n-1}{n-k-1}\frac{SSR}{TSS}$$
- Note that $\bar{R}^2 < R^2$, however if n is large the two will be very close.

Measures of fit, cont'd

- Test score example:

$$(1) \quad \text{TEST SCORE} = 698.9–2.28 \times \text{STR},$$
$$R^2 = .05, SER = 18.6$$
$$(2) \quad \text{TEST SCORE} = 686.0–1.10 \times \text{STR}–0.65 \times \text{PctEL},$$
$$R^2 = .426, \bar{R}^2 = .424, SER = 14.5$$

- What – precisely – does this tell you about the fit of regression (2) compared with regression (1)?
- Why are the R^2 and the $\bar{R}^2$ so close in (2)?

Hypothesis Testing: Single Coefficient

- Hypothesis tests and confidence intervals for a single coefficient in multiple regression follow the same logic and recipe as for the slope coefficient in a single-regressor model.
- $(\hat{\beta}_1 - \beta_1)/SE(\hat{\beta}_1)$ is approximately distributed $\mathcal{N}(0, 1)$ (CLT).
- Thus hypotheses on β_1 can be tested using the usual t-statistic, and confidence intervals are constructed as $\{\hat{\beta}_1 \pm 1.96 \times SE(\hat{\beta}_1)\}$.
- So too for $\beta_2, \ldots, \beta_k$.

Example: The California class size data

(1) $\text{TEST SCORE} = 698.9\ (10.4) – 2.28 (0.52) \times \text{STR},$

(2) $\text{TEST SCORE} = 686.0\ (8.7) – 1.10\ (0.43) \times \text{STR} – 0.65\ (0.031) \times \text{PctEL}$

- The coefficient on STR in (2) is the effect on TestScores of a unit change in STR, holding constant the percentage of English Learners in the district
- The coefficient on STR falls by one-half
- The 95% confidence interval for coefficient on STR in (2) is $\{–1.10 \pm 1.96 \times 0.43\} = (–1.95, –0.26)$
- The t-statistic testing $\beta_{STR} = 0$ is $t = –1.10/0.43 = –2.54$, so we reject the hypothesis at the 5% significance level

Application to the California Test Score in R

```
# packages to compute standard errors
library(sandwich)
library(lmtest)

library(foreign)
data = read.dta('caschool.dta')
data$score = 0.5*(data$math_scr + data$
   read_scr)
linear_model = lm(score~str+el_pct,data=
   data)

# compute standard errors, t-statistics
coeftest(linear_model, vcov. = vcovHC)
```

Table 1: Coefficients, Standard Errors, t-statistics and p-values

	Estimate	Std. Error	t-value	Pr($>|t|$)
(Intercept)	686.032244	8.812242	77.8499	$<$ 2e-16 ***
str	-1.101296	0.437066	-2.5197	0.01212 *
elpct	-0.649777	0.031297	-20.7617	$<$ 2e-16 ***

Tests of Joint Hypotheses (SW 7.2)

- Let Expn = expenditures per pupil and consider the population regression model:

$$TestScore_i = \beta_0 + \beta_1 STR_i + \beta_2 Expn_i + \beta_3 PctEL_i + u_i$$

- The null hypothesis that "school resources don't matter," and the alternative that they do, corresponds to:

$$H_0 : \beta_1 = 0 \text{ and } \beta_2 = 0$$
$$vs. H_1 : \text{ either } \beta_1 \neq 0 \text{ or } \beta_2 \neq 0 \text{ or both}$$
$$TestScore_i = \beta_0 + \beta_1 STR_i + \beta_2 Expn_i + \beta_3 PctEL_i + u_i$$

Tests of joint hypotheses, cont'd

- $H_0 : \beta_1 = 0$ and $\beta_2 = 0$ vs. $H_1 :$ either $\beta_1 \neq 0$ or $\beta_2 \neq 0$ or both
- A joint hypothesis specifies a value for two or more coefficients, that is, it imposes a restriction on two or more coefficients.
- In general, a joint hypothesis will involve q restrictions. In the example above, $q = 2$, and the two restrictions are $\beta_1 = 0$ and $\beta_2 = 0$.
- A "common sense" idea is to reject if either of the individual t-statistics exceeds 1.96 in absolute value.
- But this "one at a time" test isn't valid: the resulting test rejects too often under the null hypothesis (more than 5%)!

Why can't we just test the coefficients one at a time?

- Because the rejection rate under the null isn't 5%.
- We'll calculate the probability of incorrectly rejecting the null using the "common sense" test based on the two individual t-statistics.
- To simplify the calculation, suppose that and are independently distributed (this isn't true in general – just in this example).
- Let t_1 and t_2 be the t-statistics:

$$t_1 = \frac{\hat{\beta}_1 - 0}{SE(\hat{\beta}_1)} \text{ and } t_2 = \frac{\hat{\beta}_2 - 0}{SE(\hat{\beta}_2)}$$

- The "one at time" test is: reject $H_0 : \beta_1 = \beta_2 = 0$ if $|t_1| > 1.96$ and/or $|t_2| > 1.96$
- What is the probability that this "one at a time" test rejects H_0, when H_0 is actually true? (It should be 5%.)

Suppose t_1 and t_2 are independent (for this example)

And that the t-statistics are exactly normally distributed

The probability of incorrectly rejecting the null hypothesis using the "one at a time" test

$$
\begin{aligned}
&= \mathbb{P}(\{|t_1| > 1.96\} \cup \{|t_2| > 1.96\}) \\
&= 1 - \mathbb{P}(\{|t_1| \leq 1.96\} \cap \{|t_2| \leq 1.96\}) \\
&= 1 - \mathbb{P}(\{|t_1| \leq 1.96\}) \times \mathbb{P}(\{|t_2| \leq 1.96\}) \text{ by independence} \\
&= 1 - 0.95^2 = 0.0975 = 9.75\% > 5\%.
\end{aligned}
$$

The size of a test is the actual rejection rate under the null hypothesis.

- The size of the "common sense" test isn't 5%!
- In fact, its size depends on the correlation between t_1 and t_2 (and thus on the correlation between $\hat{\beta}_1$ and $\hat{\beta}_2$).
- **Two Solutions:**
  1. Use a different critical value in this procedure – not 1.96 (this is the "Bonferroni" method – see SW Appendix 7.1) (this method is rarely used in practice however)
  2. Use a different test statistic designed to test both β_1 and β_2 at once: the Wald statistic (this is common practice)