EE 425 (1/2015)

- **Statistical Inference in the General Linear Model (or the Multiple Regression Model)**

Statistical Inference (Hypothesis Testing)

- Testing hypothesis about an individual coefficient
- Testing the overall significance of the estimated model
- Testing the incremental contribution of an additional variable
- Testing that the coefficients satisfy certain linear restrictions
- Testing the equality or stability of parameters

The Normality Assumption of the Error Term

The distribution of test statistics:

$$\frac{\hat{\beta}_i - \beta_i}{se(\hat{\beta}_i)} \sim t_{n-k}$$

Hypothesis testing about individual regression coefficients (t test)

$\hat{CM}_t = 263.6 - 0.0056PGNP_t - 2.2316FLR_t$

se $= (11.59) \quad (0.0019) \quad (0.2099)$

$n = 64$, $R^2 = 0.7077$

How to test the effects of per capita GNP (PGNP) or female literacy rate (FLR) on child mortality rate (CM) ?

$H_0 : \beta_2 = 0 \qquad H_1 : \beta_2 < 0$

Calculate $t = -0.0056/0.009 = -2.82$

Compare with $t_{61, .05} \approx 1.671$

Can we reject H_0 ?

Testing the overall significance of the sample regression (F test)

Model: $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$

$H_0 : \beta_2 = \beta_3 = 0$ H_1: not H_0

TABLE 8.1
ANOVA Table for the Three-Variable Regression

Source of Variation	SS	df	MSS
Due to regression (ESS)	$\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i}$	2	$\frac{\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i}}{2}$
Due to residual (RSS)	$\sum \hat{u}_i^2$	$n-3$	$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-3}$
Total	$\sum y_i^2$	$n-1$	

Testing the overall significance of the sample regression (F test)

Example:

TABLE 8.3
ANOVA Table for the Child Mortality Example

Source of Variation	SS	df	MSS
Due to regression	257,362.4	2	128,681.2
Due to residuals	106,315.6	61	1742.88
Total	363,678	63	

$F = 128{,}681.2/1742.88 = 73.83$

$F_{2,61} = 19.5$

Reject H_0

F test in general

Model:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + ... + \beta_k X_{ki} + u_i$$

$H_0 : \beta_2 = \beta_3 = \beta_4 = ... = \beta_k = 0$

H_1 : Not all the slope coefficients are simultaneously zero

Test statistic:

$$F = \frac{ESS / (k-1)}{RSS / (n-k)}$$

$$F \sim F_{(k-1),(n-k)}$$

Decision rule: Reject H_0 if $F > F_{(k-1),(n-k)}$ at any chosen α

The "incremental" or "marginal" contribution of an additional explanatory variable

Old model: $Y_i = \beta_1 + \beta_2 X_{2i} + u_i$

New model: $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$

1. What is the marginal contribution of X_3 on Y when X_2 has already included in the model?
2. Is the incremental contribution of X_3 statistical significant?
3. What is the criterion for adding variables to the model?

The "incremental" contribution of an additional variable

Test statistic:

$$F = \frac{(ESS_{new} - ESS_{old}) / \text{number of new regressors}}{RSS_{new} / (n - \text{number of parameters in the new model})}$$

It can also be expressed in terms of R^2

$$F = \frac{(R^2_{new} - R^2_{old}) / \text{number of new regressors}}{(1 - R^2_{new}) / (n - \text{number of parameters in the new model})}$$

The "incremental" contribution of an additional variable

- Called Q_1 = ESS in the old model

  Q_3 = ESS in the new model

  Q_2 = ESS due to the additional regressors

  $= Q_3 - Q_1$

  Q_4 = RSS in the new model $= Q_5 - Q_3$

  Q_5 = TSS
- Suppose in the old model we include X_2 and in the new model include also X_3
- The incremental contribution of X_3 can be summarized in an ANOVA table as follows :

The "incremental" contribution of an additional variable

TABLE 8.6
ANOVA Table to Assess Incremental Contribution of a Variable(s)

Source of Variation	SS	df	MSS
ESS due to X_2 alone	$Q_1 = \hat{\beta}_{12}^2 \sum x_2^2$	1	$\frac{Q_1}{1}$
ESS due to the addition of X_3	$Q_2 = Q_3 - Q_1$	1	$\frac{Q_2}{1}$
ESS due to both X_2, X_3	$Q_3 = \hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i}$	2	$\frac{Q_3}{2}$
RSS	$Q_4 = Q_5 - Q_3$	$n-3$	$\frac{Q_4}{n-3}$
Total	$Q_5 = \sum y_i^2$	$n-1$	

The "incremental" contribution of an additional variable

Example:

1. What is the marginal contribution of female literacy rate (FLR) on CM when PGNP has already included in the model?
2. Is the incremental contribution of FLR statistical significant?
3. What is the criterion for adding variables to the model?

The "incremental" contribution of an additional variable

Estimate 2 models with n = 64

$CM\hat{}_t = 157.42 - 0.0114PGNP_t \qquad R^2 = 0.1662$

$CM\hat{}_t = 263.6 - 0.0056PGNP_t - 2.2316FLR_t$

$R^2 = 0.7077$

Test statistics in R^2 form

$$F = \frac{(0.7077 - 0.1662)/1}{(1 - 0.7077)/(64 - 3)} = 113.05$$

$F_{1,60}$ at $\alpha = 0.05 = 4$

Conclusion: Reject H_0 : The marginal effect of FLR = 0

The "incremental" contribution of an additional variable

TABLE 8.5
ANOVA Table for Regression Equation (8.4.14)

Source of Variation	SS	df	MSS
ESS (due to PGNP)	60,449.5	1	60,449.5
RSS	303,228.5	62	4890.7822
Total	363,678	63	

TABLE 8.7
ANOVA Table for the Illustrative Example: Incremental Analysis

Source of Variation	SS	df	MSS
ESS due to PGNP	60,449.5	1	60,449.5
ESS due to the addition of FLR	196,912.9	1	196,912.9
ESS due to PGNP and FLR	257,362.4	2	128,681.2
RSS	106,315.6	61	1742.8786
Total	363,678	63	

The “incremental” contribution of an additional variable

When to add a new variable in the model?

General criteria:

- When the inclusion of a variable increases the adjusted R^2
- The above criterion is equivalent to the t value of the additional variable is greater than 1 in absolute value

or

- the F- test of that incremental effect is greater than 1

Testing linear restrictions

- **Testing the equality of 2 regression coefficients, or the difference of the two coefficients equals zero.**
- **Testing the hypothesis that the sum of 2 (or several) regression coefficients are equal to a constant.**

Testing linear restrictions

Testing the Equality of Two Regression Coefficients

Model: $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + u_i$

$H_0 : \beta_3 = \beta_4$ or $H_0 : \beta_3 - \beta_4 = 0$

$H_1 : \beta_3 \neq \beta_4$ or $H_0 : \beta_3 - \beta_4 \neq 0$

Test statistics:

$$t = \frac{(\hat{\beta}_3 - \hat{\beta}_4) - (\beta_3 - \beta_4)}{se(\hat{\beta}_3 - \hat{\beta}_4)}$$

when $se(\hat{\beta}_3 - \hat{\beta}_4) = \sqrt{\mathrm{var}(\hat{\beta}_3) + \mathrm{var}(\hat{\beta}_4) - 2\,\mathrm{cov}(\hat{\beta}_3, \hat{\beta}_4)}$

Testing linear restrictions

Example:

$$\hat{Y}_i = 141.76 + 63.48X_i - 12.96X_i^2 + 0.94X_i^3$$

$$se = \qquad (4.7786) \quad (0.9857) \quad (0.0591)$$

$$\text{cov}(\hat{\beta}_3, \hat{\beta}_4) = -0.0576$$

$$t = \frac{-12.96 - 0.94}{\sqrt{(0.9857)^2 + (0.0591)^2 - 2(-0.0576)}} = -13.313$$

With n = 10, can we reject the equality of the coefficients of X^2 and X^3?

Testing linear restrictions

Testing linear equality restrictions

Suppose we want to test whether a Cobb-Douglas Production Function is constant return to scale

Cobb-Douglas Production Function:

$$Y_i = \beta_1 X_{2i}^{\beta_2} X_{3i}^{\beta_3} e^{u_i}$$

$$\ln Y_i = \ln \beta_1 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + u_i$$

H_0: $\beta_2 + \beta_3 = 1$ H_1: $\beta_2 + \beta_3 \neq 1$

There are two approaches:

- The t-test
- The F-test

Testing linear restrictions

Testing linear equality restrictions

<u>The t-test approach</u>

$H_0 : \beta_2 + \beta_3 = 1$

$H_1 : \beta_2 + \beta_3 \neq 1$

Test statistics:

$$t = \frac{(\hat{\beta}_2 + \hat{\beta}_3) - (\beta_2 + \beta_3)}{se(\hat{\beta}_2 + \hat{\beta}_3)}$$

when $se(\hat{\beta}_2 + \hat{\beta}_3) = \sqrt{\mathrm{var}(\hat{\beta}_2) + \mathrm{var}(\hat{\beta}_3) + 2\,\mathrm{cov}(\hat{\beta}_2, \hat{\beta}_3)}$

Under H_0 $t = \dfrac{(\hat{\beta}_2 + \hat{\beta}_3) - 1}{\sqrt{\mathrm{var}(\hat{\beta}_2) + \mathrm{var}(\hat{\beta}_3) + 2\,\mathrm{cov}(\hat{\beta}_2, \hat{\beta}_3)}}$

Testing linear restrictions

Testing linear equality restrictions

The F-test approach

Estimate the restricted and unrestricted model to obtain the test statistic

$$F = \frac{(RSS_R - RSS_{UR})/m}{RSS/(n-k)} = \frac{(\sum \hat{u}_R^2 - \sum \hat{u}_{UR}^2)/m}{\sum \hat{u}_{UR}^2/(n-k)}$$

In the form of R^2

$$F = \frac{(R_{UR}^2 - R_R^2)/m}{(1-R_{UR}^2)/(n-k)}$$

m is the number of restriction, n = sample size

k = number of parameters estimated in the unrestricted model

Testing linear restrictions

The restricted model to test constant return to scale

Cobb-Douglas Production Function:

$$\ln Y_i = \ln \beta_1 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + u_i$$

$$H_0 : \beta_2 + \beta_3 = 1 \quad \text{or} \quad \beta_2 = 1 - \beta_3$$

Hence the restricted model

$$\ln Y_i = \ln \beta_1 + (1 - \beta_3) \ln X_{2i} + \beta_3 \ln X_{3i} + u_i$$

$$(\ln Y_i - \ln X_{2i}) = \beta_1 + \beta_3 (\ln X_{3i} - \ln X_{2i}) + u_i$$

$$\ln\left(\frac{Y_i}{X_{2i}}\right) = \beta_1 + \beta_3 \ln\left(\frac{X_{3i}}{X_{2i}}\right) + u_i$$

Testing linear restrictions

Example of F- test

Unrestricted model:

$\ln(GDP_t) = -1.65 + 0.34\ln(Labor_t) + 0.846\ln(Capital_t)$

$R^2_{UR} = 0.9954$ $RSS_{UR} = 0.0136$

Restricted model:

$\ln(GDP/Labor)_t = -0.4947 + 1.0153\ln(Capital/Labor)_t$

$R^2_{R} = 0.9777$ $RSS_{R} = 0.0166$

$n = 20$

$$F = \frac{(RSS_R - RSS_{UR})/m}{RSS/(n-k)} = \frac{(0.0166 - 0.0136)/1}{(1-0.0136)/(20-3)} = 3.75$$

At $\alpha = .05$, $F_{1,17} = 4.45$, Conclusion: cannot reject H_0

Testing linear restrictions

General F- test

All the tests mentioned so far could be express as some linear restriction in the parameter, hence F-test can apply to all cases

Example 8.4 on pg. 253 in Gujarati

$$\ln Y_i = \beta_1 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + \beta_4 \ln X_{4i} + \beta_5 \ln X_{5i} + u_i$$

Y_i = per capita consumtpion of chicken

X_{2i} = real disposable per capita income

X_{3i} = real retail price of chicken per lb.

X_{4i} = real retail price of pork per lb.

X_{5i} = real retail price of beef per lb.

$H_0 : \beta_4 = \beta_5 = 0 \qquad H_1 :$ Not both β_4 and $\beta_5 = 0$

Testing linear restrictions

General F- test

To proceed:

Step 1: Estimate the unrestricted model

$$\ln \hat{Y}_i = 2.1898 + 0.3424 \ln X_{2i} - 0.5046 \ln X_{3i} + 0.1485 \ln X_{4i} + 0.0911 \ln X_{5i}$$

$$R^2_{UR} = 0.9823$$

Step 2: Estimate the restricted model

$$\ln \hat{Y}_i = 2.0328 + 0.4515 \ln X_{2i} - 0.3772 \ln X_{3i}$$

$$R^2_R = 0.9801$$

Test statistic:

$$F = \frac{(R^2_{UR} - R^2_R)/m}{(1 - R^2_{UR})/(n-k)}$$

$$= \frac{(0.9823 - 0.9801)/2}{(1 - 0.9823)/18} = 1.1224$$

$$F_{2,18} = 3.55 \text{ at } \alpha = 0.05$$

Testing for Equality or Stability of Two Regression Models: The Chow test

- To test whether there is change in the relationship between the dependent and the independent variable
- Example: 1970-1981, 1982-1995 = pre- and post-recession,

TABLE 8.9
Savings and Personal Disposable Income (billions of dollars), United States, 1970–1995

Source: *Economic Report of the President*, 1997, Table B-28, p. 332.

Observation	Savings	Income	Observation	Savings	Income
1970	61.0	727.1	1983	167.0	2522.4
1971	68.6	790.2	1984	235.7	2810.0
1972	63.6	855.3	1985	206.2	3002.0
1973	89.6	965.0	1986	196.5	3187.6
1974	97.6	1054.2	1987	168.4	3363.1
1975	104.4	1159.2	1988	189.1	3640.8
1976	96.4	1273.0	1989	187.8	3894.5
1977	92.5	1401.4	1990	208.7	4166.8
1978	112.6	1580.1	1991	246.4	4343.7
1979	130.1	1769.5	1992	272.6	4613.7
1980	161.8	1973.3	1993	214.4	4790.2
1981	199.1	2200.2	1994	189.4	5021.7
1982	205.5	2347.3	1995	249.3	5320.8

Testing for Equality or Stability of Two Regression Models: The Chow test

Estimate 3 regressions:

(1) 1970-1981: $Y_t = \lambda_1 + \lambda_2 X_t + u_{1t}$ $\quad n_1 = 12$

(2) 1982-1995: $Y_t = \gamma_1 + \gamma_2 X_t + u_{2t}$ $\quad n_2 = 14$

(3) 1970-1995: $Y_t = \alpha_1 + \alpha_2 X_t + u_t$ $\quad n = (n_1 + n_2) = 26$

If there is no structural change, regression # 3 should be used, otherwise, we should use regression # 1 and # 2

Testing for Equality or Stability of Two Regression Models: The Chow test

Assumptions in Chow Test

1) $u_{1t} \sim N(0,\sigma^2)$ and $u_{2t} \sim N(0,\sigma^2)$

i.e. variances of the error terms for 2 sub-periods are equal

2) u_{1t} and u_{2t} are independently sidtributed.

Steps to the Test

1) Estimate the regression when there is no structural change with n_1+n_2 samples, obtained RSS_R with n_1+n_2-k df.
2) Estimate 2 regressions under structural change, obtain RSS_1 with n_1-k df. and RSS_2 with n_2-k df.

Testing for Equality or Stability of Two Regression Models: The Chow test

Steps to the Test (cont.)

3) Define $RSS_{UR} = RSS_1 + RSS_2$ as the unrestricted residual sum of squares with (n_1+n_2-2k df.)

4) The test statistics

$$F = \frac{(RSS_R - RSS_{UR})/k}{RSS_{UR}/(n_1+n_2-2k)} \sim F_{k,(n_1+n_2-2k)}$$

Testing for Equality or Stability of Two Regression Models: The Chow test

Example

1) $\hat{Y}_t = 1.0161 + 0.0803X_t$

   $R^2 = 0.9021 \quad RSS_1 = 1785.032 \quad df = 10$

2) $\hat{Y}_t = 153.4947 + 0.0148X_t$

   $R^2 = 0.2971 \quad RSS_2 = 10,005.22 \quad df = 12$

3) $\hat{Y}_t = 62.4226 + 0.0376X_t$

   $R^2 = 0.7672 \quad RSS_R = 23,248.30 \quad df = 24$

$$F = \frac{(23,248.30 - 11,790.252)/2}{11,790.252/22} = 10.69$$

$F_{2,22} = 5.72$ at $\alpha = .01$

Hence reject that there is no structural change

Testing for Equality or Stability of Two Regression Models: The Chow test

To test whether the variances of the error terms in two sub-period are equal.

$H_0: \ \sigma_1^2 = \sigma_2^2 \quad H_0: \ \sigma_1^2 \neq \sigma_2^2$

Test statistic

$$F = \frac{(\hat{\sigma}_1^2 / \sigma_1^2)}{(\hat{\sigma}_2^2 / \sigma_2^2)} \sim F_{(n_1-k),(n_2-k)}$$

Under H_0 $F = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2}$

By convention we always put the larger estimate in the numerator

Testing for Equality or Stability of Two Regression Models: The Chow test

$$\text{Since } \hat{\sigma}_1^2 = \frac{RSS_1}{n_1 - k} = 1785.032/10 = 178.5032$$

$$\text{and } \hat{\sigma}_2^2 = \frac{RSS_2}{n_2 - k} = 10{,}005.22/12 = 833.7693$$

$$F = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} = \frac{833.7693}{178.5032} = 4.6701$$

$$F_{12,10} = 2.91 \text{ with } \alpha = .05$$