ECONOMETRICS-I Lecture Notes | Chapter 9

The Chow Test

The Chow Test is used to test if there is a structural change in the Regression Parameters at a specific year. Test is performed as follows:

Example: Y: Savings ; X: Income

(I) Time Period: 1970-1981: $Y_t = \lambda_1 + \lambda_2 X_t + u_{1t}$ $\quad n=12$; RSS_1

(II) Time Period: 1982-1995: $Y_t = \gamma_1 + \gamma_2 X_t + u_{2t}$ $\quad n=14$; RSS_2

(III) Time Period: 1970-1995 $Y_t = \alpha_1 + \alpha_2 X_t + u_t$ $\quad n = 12+14 = 26$; RSS_3

Obtain; $\boxed{RSS_{UR} = RSS_1 + RSS_2}$ and $\boxed{RSS_R = RSS_3}$

Is there a Structural Change in 1982?

We assume, the two time periods residual terms are;

$$u_{1t} \sim Normal(0; \sigma^2)$$
$$u_{2t} \sim Normal(0; \sigma^2)$$

Before making the Chow test, we test if equality of Variances assumption is valid;

(i) $H_0: \sigma_1^2 = \sigma_2^2$
$\quad H_A: \sigma_1^2 \neq \sigma_2^2$

(ii) $F = \dfrac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} = \dfrac{RSS_1/(n_1-k)}{RSS_2/(n_2-k)}$ → Larger Variance

And the Chow test is;

(i) $H_0: \lambda_1 = \gamma_1 (=\alpha_1)$ // $\lambda_2 = \gamma_2 (=\alpha_2)$: No structural change

H_A: At least one restriction is NOT valid

(ii) $F = \dfrac{(RSS_R - RSS_{UR})/k}{RSS_{UR}/(n_1+n_2-2k)}$

Alp Giray Özen | 0533 549 91 08 | alp@lecturemania.com | www.lecturemania.com

Consider the following estimated models:

(I) 1970 - 1981: $\hat{y}_t = 1{,}0161 + 0{,}0803 X_t$ $\quad (1981-1970)+1$

$t = (0{,}0873) \quad (9{,}6015)$

$R_1^2 = 0{,}9021 \quad RSS_1 = 1785 \quad df = n_1 - 2 = 12 - 2 = 10$

(II) 1982 - 1995: $\hat{y}_t = 153{,}4947 + 0{,}0148 X_t$ $\quad (1995-1982)+1$

$t = (4{,}6922) \quad (1{,}7707)$

$R_2^2 = 0{,}2071 \quad RSS_2 = 10005 \quad df = n_2 - 2 = 14 - 2 = 12$

(III) 1970 - 1995: $\hat{y}_t = 62{,}4226 + 0{,}0376 X_t$

$t = (4{,}8917) \quad (8{,}8973)$

$R_3^2 = 0{,}7672 \quad RSS_3 = 23248 \quad df = n_1 + n_2 - 2 = 12 + 14 - 2 = 24$

Test of Equality of Variances:

$$\hat{\sigma}_1^2 = \frac{RSS_1}{n_1 - 2} = \frac{1785}{10} = 178{,}5 \; ; \quad \hat{\sigma}_2^2 = \frac{RSS_2}{n_2 - 2} = \frac{10005}{12} = 833{,}8$$

(i) $H_0: \sigma_1^2 = \sigma_2^2$

$H_A: \sigma_1^2 \neq \sigma_2^2$

$\alpha = 0{,}10$

(ii) $F = \dfrac{\hat{\sigma}_2^2 \rightarrow df = 12}{\hat{\sigma}_1^2 \rightarrow df = 10}$

(iii)

Reject H_0 if $F > 2{,}91$

(iv) $F = \dfrac{833{,}8}{178{,}5} = 4{,}67$

(v) Reject H_0.

In fact, since $\sigma_1^2 = \sigma_2^2$ assumption is NOT valid, Chow-Test is NOT applicable. However, to follow the steps, we'll apply it.

(i) $H_0: \lambda_1 = \gamma_1$ AND $\lambda_2 = \gamma_2$

H_A: At least one restriction is NOT valid

$\alpha = 0{,}05$

(ii) $F = \dfrac{(RSS_R - RSS_{UR})/k \rightarrow 2}{RSS_{UR}/(n_1 + n_2 - 2k) \rightarrow 22}$

(iii)

Reject H_0 if $F > 5{,}72$

(iv) $RSS_R = 23248$

$RSS_{UR} = 1785 + 10005 = 11790$

$$F = \frac{(23248 - 11790)/2}{11790/22} = 10{,}69$$

(v) Reject H_0.

There is a structural change

Dummy VARIABLES

Dummy Variable D_i is a binary variable;

$$D_i = \begin{cases} 1 & \text{if unit is in the category} \\ 0 & \text{otherwise} \end{cases}$$

We use Dummy Variables to model categorical (qualitative) variables.

Example Let Y: Food Expenditure

$$D = \begin{cases} 1 & \text{if person is female} \\ 0 & \text{o.w. (male)} \end{cases}$$

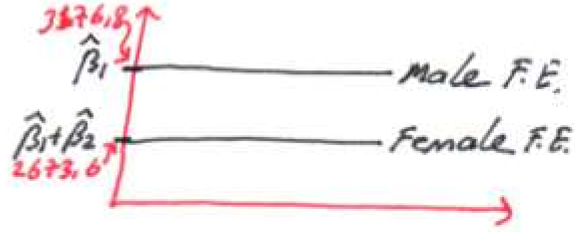
$$Y_i = \beta_1 + \beta_2 \cdot D_i + u_i$$

When we estimate this model; $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 D_i$

Estimated food expenditure of a female is;

$(\hat{Y}_i \mid D_i = 1) = \hat{\beta}_1 + \hat{\beta}_2$ and a male is;

$(\hat{Y}_i \mid D_i = 0) = \hat{\beta}_1$



Consider the estimated Model;

$$\hat{Y} = 3176,800 - 503,2\, D_i$$
$$t = \quad (13,63) \qquad (-2,53) \qquad R^2 = 0,18$$

Female $\Rightarrow (\hat{Y}_i \mid D_i = 1) = 3176,8 - 503,2 = 2673,6$

Male $\Rightarrow (\hat{Y}_i \mid D_i = 0) = 3176,8$

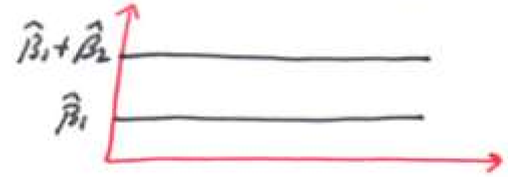
So, $\hat{\beta}_2$ is interpreted as "Contribution of being a female"

Then, on the average, Females have 503,2 less food expenditure than that of Males.

β_1 is "base category": when all dummies are zero.

If we rewrite the regression for $D_i = \begin{cases} 1 & \text{male} \\ 0 & \text{female} \end{cases}$

We have; $\hat{Y}_i = \underbrace{2673,6}_{=\hat{\beta}_1} + \underbrace{503,2}_{=\hat{\beta}_2} D_i$



We can use Dummy Variables and Quantitative variables in the same regression. Consider the model,

$$Y_i = \beta_1 + \beta_2 X_i + \alpha D_i + u_i$$

where; Y_i and D_i are same ($D_i = 1$ for female) and X: Yearly Income. Given income, the estimated food expenditures for a Female and a Male are;

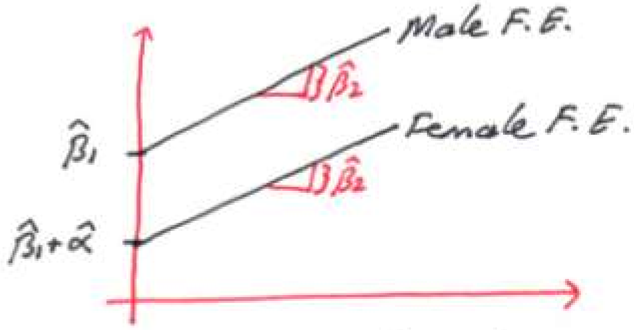
$$(\hat{Y}_i | D_i = 1) = \hat{\beta}_1 + \alpha + \hat{\beta}_2 X_i \quad \text{and} \quad (\hat{Y}_i | D_i = 0) = \hat{\beta}_1 + \hat{\beta}_2 X_i$$

Let, the estimated model is;

$$\hat{Y} = 1506,2 - 228,9\, D_i + 0,06 X_i$$
$$t = (8,01) \quad (-2,14) \quad (9,64) \qquad R^2 = 0,92$$

If, for example, income is 15000 $;

$$(\hat{Y}_i | D_i = 1, X_i = 15000) = 1506,2 - 228.9 + 0,06.15000 = 2177,1$$
$$(\hat{Y}_i | D_i = 0, X_i = 15000) = 1506,2 + 0,06 . 15000 = 2406$$



Again, we can rewrite the regression for $D_i = \begin{cases} 1 & \text{male} \\ 0 & \text{female} \end{cases}$ as;

$$\hat{Y} = 1277,3 + 228,9 . D_i + 0,06 X_i$$

The following Specifications are wrong!

(i)

$$E = \begin{cases} 0 & 2 \text{ years} \\ 1 & 4 \text{ years} \\ 2 & \text{Master} \end{cases} \qquad y_i = \beta_1 + \alpha . E_i + u$$

E is NOT Binary Here.

(ii) $E_1 = \begin{cases} 1 & \text{if 2 years} \\ 0 & \text{o.w.} \end{cases}$ $E_2 = \begin{cases} 1 & \text{if 4 years} \\ 0 & \text{o.w.} \end{cases}$ $E_3 = \begin{cases} 1 & \text{if Master} \\ 0 & \text{o.w.} \end{cases}$

$$y_i = \beta_1 + \alpha_1 . E_{1i} + \alpha_2 . E_{2i} + \alpha_3 . E_{3i} + u$$

This is also NOT true because Base: β_1 has NO meaning. This is called "Dummy Variable Trap"

To overcome Dummy Variable Trap, we can do one of the following two:

(i) Drop Constant β_1:

$$y_i = \alpha_1 E_{1i} + \alpha_2 E_{2i} + \alpha_3 E_{3i}$$

Then, data units must fall one of these: $\{(1,0,0)\ (0,1,0)\ (0,0,1)\}$
Here, there is NO Base Category

(ii) Drop one of the Dummies so that dropped dummy become a base. For example, if we drop E_3, then $(0,0)$ would represent "Master degree", whose mean is Base: β_1

$$y_i = \beta_1 + \alpha_1 E_{1i} + \alpha_2 E_{2i} + u_i$$

Consider the estimated Regression;

$$\hat{y} = 1800 - 500 E_1 - 200 E_2$$

$$t = \quad (17{,}34) \quad (-3{,}61) \quad (-2{,}56) \qquad R^2 = 0{,}3147 \quad n = 169$$

Example: Consider the following SRF:

$$\hat{y}_i = 8{,}81 + 1{,}1\, D_{2i} - 1{,}67\, D_{3i}$$
$$t = (21{,}95) \quad (2{,}37) \quad (-3{,}44) \quad n = 528,\ R^2 = 0{,}0332$$

where, y = Hourly Wage ; $D_2 = \begin{cases} 1 & \text{married} \\ 0 & \text{o.w.} \end{cases}$ D_3 = Region $\begin{cases} 1 & \text{South} \\ 0 & \text{o.w.} \end{cases}$

what is the Base (Benchmark) Category here? Obviously, it is unmaried, Non-South residence. The mean hourly wage of Base category is 8,81\$.

A maried person, on the average, is expected to earn 1,1\$ more. Then, Mean hourly wage of married, Non-South residence is $8{,}81 + 1{,}1 = 9{,}91$ \$.

A South Resident, on the average, is expected to earn 1,67 less than Base category, which is $8{,}81 - 1{,}67 = 7{,}14$.

If we let $E(y) = \beta_0 + \alpha_1 D_{2i} + \alpha_2 D_{3i}$

Note that both α_1 and α_2 are significant.

Dummy Variable TRAP;

So far, our categorical variables had two levels. How can we construct a regression model with a categorical variable that has more than two levels?

let, we want to estimate Teacher's salary based on his/her education where; y = Salary of the teacher

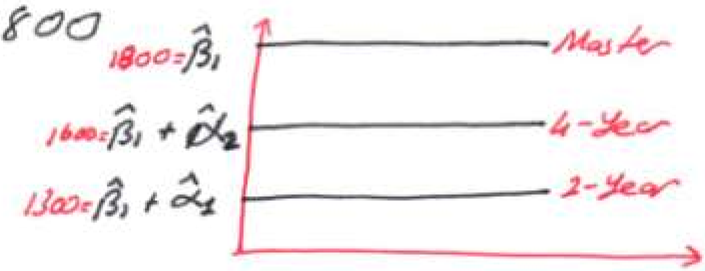
E = Education = $\begin{cases} \text{2 years vacational School} \\ \text{4 years University} \\ \text{Master Degree.} \end{cases}$

We have;

2 Year Ed: $(\hat{y} \mid E_1 = 1, E_2 = 0) = 1800 - 500 = 1300$

4 Year Ed: $(\hat{y} \mid E_1 = 0, E_2 = 1) = 1800 - 200 = 1600$

Master: $(\hat{y} \mid E_1 = 0, E_2 = 0) = 1800$



If we drop E_1, We have

$$\hat{y}_i = 1300 + 300 E_{2i} + 500 E_{3i}$$

where 2-Year Education become the base category.

If we drop β_1, we have

$$\hat{y}_i = 1300 E_{1i} + 1600 E_{2i} + 1800 E_{3i}$$

where the model has NO Base.

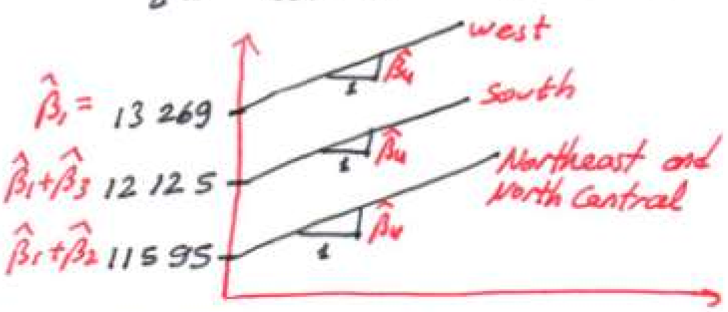
Example// Consider the following model

$$y_i = \beta_1 + \beta_2 D_{2i} + \beta_3 D_{3i} + \beta_4 X_i + u_i$$

Where: y_i = Average annual Salary ; X_i = Spending on public school per pupil($)

$D_{2i} = \begin{cases} 1 & \text{if Northeast or North Central} \\ 0 & \text{o.w.} \end{cases}$ $\quad D_{3i} = \begin{cases} 1 & \text{if South} \\ 0 & \text{o.w.} \end{cases}$

$$\hat{y}_i = 13269{,}1 - 1673{,}5 D_{2i} - 1144{,}2 D_{3i} + 3{,}3 X_i$$
$$t = \quad (9{,}51) \quad\quad (-2{,}09) \quad\quad (-1{,}33) \quad\quad (10{,}35)$$



9.2. Consider the following regression results (t ratios are in parentheses)*:

$$\hat{Y}_i = \underset{(4.67)}{1286} + \underset{(3.70)}{104.97X_{2i}} - \underset{(-3.80)}{0.026X_{3i}} + \underset{(0.24)}{1.20X_{4i}} + \underset{(0.08)}{0.69X_{5i}}$$

$$\underset{(-0.40)}{-19.47X_{6i}} + \underset{(6.94)}{266.06X_{7i}} - \underset{(-3.04)}{118.64X_{8i}} - \underset{(-6.14)}{110.61X_{9i}}$$

$$R^2 = 0.383 \qquad n = 1543$$

where Y = wife's annual desired hours of work, calculated as usual hours of work per year plus weeks looking for work
X_2 = after-tax real average hourly earnings of wife
X_3 = husband's previous year after-tax real annual earnings
X_4 = wife's age in years
X_5 = years of schooling completed by wife
X_6 = attitude variable, 1 = if respondent felt that it was all right for a woman to work if she desired and her husband agrees, 0 = otherwise
X_7 = attitude variable, 1 = if the respondent's husband favored his wife's working, 0 = otherwise
X_8 = number of children less than 6 years of age
X_9 = number of children in age groups 6 to 13

a. Do the signs of the coefficients of the various nondummy regressors make economic sense? Justify your answer.
b. How would you interpret the dummy variables, X_6 and X_7? Are these dummies statistically significant? Since the sample is quite large, you may use the "2-t" rule of thumb to answer the question.
c. Why do you think that age and education variables are not significant factors in a woman's labor force participation decision in this study?

9.2) a) The expected Signs are;

✓ $\hat{\beta}_2 > 0$: Higher Wage ⇒ More Work

✓ $\hat{\beta}_3 < 0$: Husband Earns More ⇒ less desire of working hours for wife

? $\hat{\beta}_4$? : Depends on age interval of the study and culture

✓ $\hat{\beta}_5 > 0$: More education ⇒ More opportunity to work

✗ $\hat{\beta}_6 > 0$ if respondent feels working is all right, expected sign of working hours is positive. Interesting result!

✓ $\hat{\beta}_7 > 0$ if husband favors working, more working hours is expected.

✓ $\hat{\beta}_8 < 0$
✓ $\hat{\beta}_9 < 0$ } more children ⇒ less oportuniti or time to work.

b) $\hat{\beta}_6$: If the respondent feels it is all right a woman works, on the average, desired hours of work is 19,47 less.

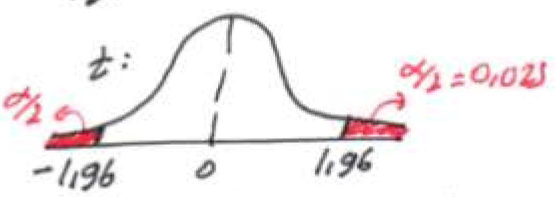
$\hat{\beta}_7$: If the respondent's husband favors his wife working, on the average, desired hours of work is 266,06 more.

$H_0: \beta_6 = 0$ | $H_0: \beta_7 = 0$

$H_A: \beta_6 \neq 0$ | $H_A: \beta_7 \neq 0$

$\alpha = 0{,}05$ | $\alpha = 0{,}05$

$$t = \frac{\hat{\beta}_j - \beta_j}{SE(\hat{\beta}_j)}\ ;\ df = n - 9 = 1543 - 9 = 1534$$



Reject H_0 if $|t| > 1{,}96$

$t = -0{,}40$ | $t = 6{,}94$

Do NOT Reject H_0 | Reject H_0

β_6 is NOT significant and β_7 is significant. (Then, sign of $\hat{\beta}_6$ is NOT important.)

c) Age depends on age interval of the study and culture, so Age variable being NOT significant is a usual result. Education is expected to be a significant factor. Maybe, for a woman, other factors shadows the significance of education, such that Number of Children, Husband's Permision or Husband's Wage.

The Dummy Variable Alternative to the Chow Test

Consider the structural Change Question again:

1970-1981: $y = \lambda_1 + \lambda_2 X_t + u_{1t}$ $\qquad$ y: Savings

1982-1995: $y_t = \gamma_1 + \gamma_2 X_t + u_{2t}$ $\qquad$ X: Income

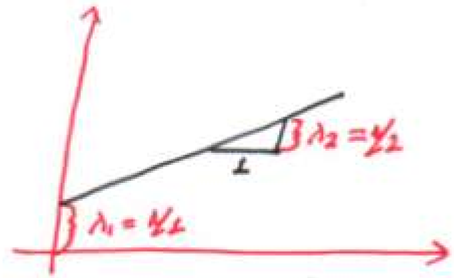
Let, instead of simultaneously testing $\lambda_1 = \gamma_1$ and $\lambda_2 = \gamma_2$, we have two different hypothesis testing:

$H_{01}: \lambda_1 = \gamma_1$ $\qquad$ $H_{02}: \lambda_2 = \gamma_2$

$H_{A1}: \lambda_1 \neq \gamma_1$ $\qquad$ $H_{A2}: \lambda_2 \neq \gamma_2$

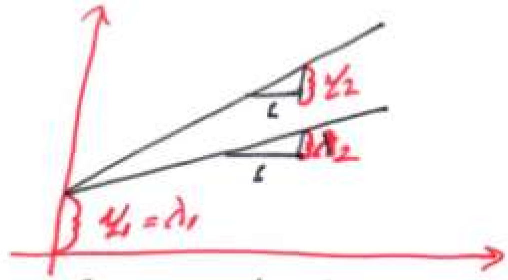
For the two decisions of tests, we have 4 cases:

(i) Do NOT Reject H_{01} and Do NOT Reject H_{02}



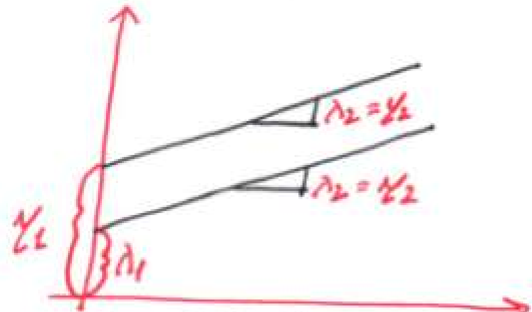
Coincident Regressions

(iii) Do NOT Reject H_{01} and Reject H_{02}



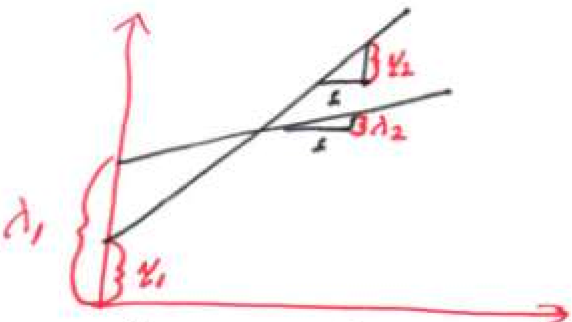
Concurrent Regressions

(ii) Reject H_{01} and Do NOT Reject H_{02}



Parallel Regressions

(iv) Reject H_{01} and Reject H_{02}



Dissimilar Regressions

Now, Consider the Model; $D_t = \begin{cases} 1 & \text{for observations in 1982-1995} \\ 0 & \text{o.w.} \end{cases}$

$$Y_t = \alpha_1 + \alpha_2 D_t + \beta_1 X_t + \beta_2 (D_t X_t) + u_t$$

Mean Saving for 1970-1981

$$E(Y_t \mid D_t = 0, X_t) = \alpha_1 + \beta_1 X_t$$

Mean Saving for 1982-1995

$$E(Y_t \mid D_t = 1, X_t) = (\alpha_1 + \alpha_2) + (\beta_1 + \beta_2) X_t$$

Then; $H_{01}: \alpha_2 = 0$ $\qquad$ $H_{02}: \beta_2 = 0$

$H_{A1}: \alpha_2 \neq 0$ $\qquad$ $H_{A2}: \beta_2 \neq 0$

Gives the same ideas we have explained previously.

Example;

$$\hat{Y}_t = 1,0161 + 152,4786 D_t + 0,0803 X_t - 0,0655 (D_t X_t)$$
$$t = (0,0504) \quad (4,61) \quad (5,54) \quad (-4,10)$$

From this regression, we can write the equations (I) and (II) as we wrote in the Chow Test;

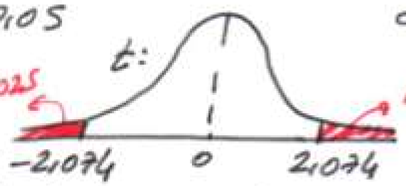
$$(\hat{Y}_t \mid D_t = 0) = 1,0161 + 0,0803 X_t$$

$$(\hat{Y}_t \mid D_t = 1) = (1,061 + 152,4786) + (0,0803 - 0,0655) X_t$$
$$= 153,4947 + 0,0148 X_t$$

To test the structural change (and the kind of change), we test: $H_{01}: \alpha_2 = 0$ $\qquad$ $H_{02}: \beta_2 = 0$

$H_{A1}: \alpha_2 \neq 0$ $\qquad$ $H_{A2}: \beta_2 \neq 0$

$\alpha = 0,05$ $\qquad$ $\alpha = 0,05$



$df = n - 4 = 26 - 4 = 22$

Note that both α_2 and β_2 are significant. The periods have dissimilar regressions.

9.3. Consider the following regression results.* (The actual data are in Table 9.7.)

$$\widehat{\text{UN}}_t = 2.7491 + 1.1507D_t - 1.5294V_t - 0.8511(D_tV_t)$$
$$t = (26.896) \quad (3.6288) \quad (-12.5552) \quad (-1.9819)$$
$$n=50 \qquad R^2 = 0.9128$$

where UN = unemployment rate, %
V = job vacancy rate, %
D = 1, for period beginning in 1966–IV
= 0, for period before 1966–IV
t = time, measured in quarters

Note: In the fourth quarter of 1966, the then Labor government liberalized the National Insurance Act by replacing the flat-rate system of short-term unemployment benefits by a mixed system of flat-rate and (previous) earnings-related benefits, which increased the level of unemployment benefits.

a. What are your prior expectations about the relationship between the unemployment and vacancy rates?

b. Holding the job vacancy rate constant, what is the average unemployment rate in the period beginning in the fourth quarter of 1966? Is it statistically different from the period before 1966 fourth quarter? How do you know?

c. Are the slopes in the pre- and post-1966 fourth quarter statistically different? How do you know?

d. Is it safe to conclude from this study that generous unemployment benefits lead to higher unemployment rates? Does this make economic sense?

9.3) a) Since they are conversely related, the expected signs of slope coefficients $\hat{\beta}_3$ and $\hat{\beta}_4$ are negative

b) $(\widehat{UN}_t \mid V = 0{,}819; D = 1) = 2{,}75 + 1{,}15 - 1{,}53 \cdot 0{,}819 - 0{,}85 \cdot 0{,}819 = 1{,}9$

$$UN_t = \beta_1 + \beta_2 D_t + \beta_3 V_t + \beta_4 (D_t V_t) + u_t$$

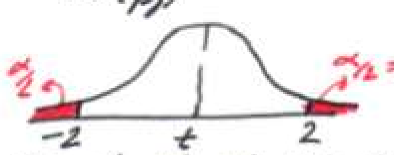
(i) $H_{01}: \beta_2 = 0$ $\qquad H_{02}: \beta_4 = 0$

$H_{A1}: \beta_2 \neq 0$ $\qquad H_{A2}: \beta_4 \neq 0$

$\alpha = 0{,}05$ $\qquad \alpha = 0{,}05$

(ii) $t = \dfrac{\hat{\beta}_j - \beta_j}{SE(\hat{\beta}_j)}$; $df = n - 4 = 46$

(iii)



Reject H_0 if $|t| > 2$

(iv) Reject H_0 $\qquad$ Do NOT Reject H_0

$t = 3{,}63$ $\qquad t = 1{,}98$

(v) There's a structural change by intercept, Parallel Regressions.

Yes, a significant shift has occured at period 1966–IV.

c) β_3 is significant since $|-12,55| > 2$. $(\beta_3+\beta_4)$ is also significant because it is more negative than β_3. So, Both pre and post-1966-IV quarter slopes are significant.

d) Yes, because there is a significant shift of 1,15 after increasing the level of unemployement benefits.

9.16. To study the rate of growth of population in Belize over the period 1970–1992, Mukherjee et al. estimated the following models[†]:

$$\text{Model I:} \quad \widehat{\ln(\text{Pop})_t} = 4.73 + 0.024t$$
$$t = (781.25) \quad (54.71)$$

Growth rate: 2,4%

$$\text{Model II:} \quad \widehat{\ln(\text{Pop})_t} = 4.77 + 0.015t - 0.075D_t + 0.011(D_t t)$$
$$t = (2477.92) \quad (34.01) \quad (-17.03) \quad (25.54)$$

where Pop = population in millions, t = trend variable, $D_t = 1$ for observations beginning in 1978 and 0 before 1978, and ln stands for natural logarithm.

a. In Model I, what is the rate of growth of Belize's population over the sample period?

b. Are the population growth rates statistically different pre- and post-1978? How do you know? If they are different, what are the growth rates for 1972–1977 and 1978–1992?

9.16) a) Remember, in log-lin model, $100 \cdot \hat{\beta}_2$ gives the estimated growth rate since it is expected effect of a unit ~~change~~ increase in t to a percentage increase in y.

Estimated rate of growth is 2,4%.

b) Yes, because the increase in growthrate is β_4 and it is significantly greater than 0. ($t = 25,54$, very high significance).

1972-1977: $(\widehat{\ln Pop_t} \mid D_t = 0) = 4,77 + 0,015t$ → Growth Rate: 1,5%

1978-1992: $(\widehat{\ln Pop_t} \mid D_t = 1) = (4,77 - 0,075) + (0,015 + 0,011)t$
$= 4,695 + 0,026t$ → Growth Rate: 2,6%

Dummy Variables in Seasonal Analysis:

Consider the model;

$$Y_t = \alpha_1 D_{1t} + \alpha_2 D_{2t} + \alpha_3 D_{3t} + \alpha_4 D_{4t} + u_t$$

where; Y_t = Sales of Refregirators and D's are dummies taking a value of 1 in the relevant quarter and 0 o.w.

$$\hat{Y}_t = 1222,13\, D_{1t} + 1467,50 D_{2t} + 1569,75 D_{3t} + 1160,00 D_{4t}$$
$$t = \quad (20,37) \qquad (24,46) \qquad (26,17) \qquad (19,36)$$

Here, estimated α coefficients represent the average, or mean, sales of refregirators in each season (i.e. quarter)

If we drop D_{1t} from the model, the model becomes

$$\hat{Y}_t = 1222,13 + \underbrace{245,38}_{=1467,5-1222,13} D_{2t} + \underbrace{347,63}_{=1569,75-1222,13} D_{3t} \underbrace{-62,13}_{=1160-1222,13} D_{4t}$$
$$t = \quad (20,37) \qquad (2,89) \qquad (4,00) \qquad (-0,73)$$

Note that, α_2 and α_3 are significant which means Sales for 2nd and 3rd quarters are significantly different from the first quarter. However, α_4 is NOT significant. Sales for 1st and 4th quarters are NOT significantly different.

To practice, we may drop another dummy from the model, let's say D_3. The model becomes;

$$\hat{Y}_t = 1569,75 - 347,62 D_{1t} - 102,25 D_{2t} - 409,75 D_{4t}$$

2. A real estate economist analyzing the factors those contribute to the price changes in the real estate market. He collected data on two similar neighbourhoods, one bordering a large university and one that is a neighbourhood about 10 kilometres from the university. Data consists of 1000 observations and is reported in Table9-1.xls. The data file consists of the following variables:

$PRICE_t$ = House prices given in $'s

$SQFT$ = Number of square feet of living area,

AGE_t = Age of the house,

$UTOWN_t = 1$ for houses near the university, 0 otherwise,

$POOL_t = 1$ for houses with a pool, 0 otherwise,

$FPLACE_t = 1$ for houses with a fireplace, 0 otherwise.

The economist specifies the regression equation as:

$$PRICE_t = \beta_1 + \beta_2 SQFT_t + \delta_1 UTOWN_t + \lambda(SQFT_t * UTOWN_t) + \beta_3 AGE_t + \delta_2 POOL_t + \delta_3 FPLACE_t + u_t$$

a) State the expected signs of the coefficients and give economic reasons for these signs.
b) What is the base regression and which group of houses does it describe?
c) What is the estimated marginal effect of size (SQFT) on the price of the house?
d) Test the hypothesis that this marginal effect depends on the location of the house.
e) By how much do the houses depreciate per year? Is this statistically significant?
f) Test the hypothesis that fireplace has a positive impact on prices.
g) Test the hypothesis that pool has a positive impact on prices.
h) Test the hypothesis that presence of pool contributes more to price than the presence of fireplace.
i) Test the hypothesis that location has no impact on houses.
j) Rewrite a new regression equation which captures the differential effect of depreciation for houses with a pool and without a pool.

Table 9.2 House Price Equation Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	24500	6191.721	3.957	0.0001
UTOWN	1	27453	8422.582	3.259	0.0012
SQFT	1	76.121766	2.451	31.048	0.0001
USQFT	1	12.994049	3.320	3.913	0.0001
AGE	1	-190.086422	51.204	-3.712	0.0002
POOL	1	4377.163290	1196.691	3.658	0.0003
FPLACE	1	1649.175634	971.956	1.697	0.0901

On this final example, we'll review the concepts on Dummy Variables and we'll learn interaction effect

a) Estimated signs are;

$\hat{\beta}_3 < 0$ and all other $\hat{\beta}_j > 0$. The reason is that; Price of a house increases by SQFT and by being at UTOWN and by having POOL and FPLACE. The only negative effect on price of a house is its AGE. (Older house is cheaper)

The estimated model is;

$$\widehat{PRICE} = 24500 + 76{,}12\, SQFT + 27453\, UTOWN + 12{,}99\,(SQFT * UTOWN)$$
p-value (0,0001) (0,0001) (0,0012) (0,0001)

$$-190{,}09 \cdot AGE + 4377{,}16\, POOL + 1649{,}18 \cdot FPLACE$$
(0,0002) (0,0003) (0,0901)

b) Base Regression is the regression at which all dummies are 0. It describes the price of the house that is NOT in the UTOWN, has NO POOL and NO FPLACE. That is; Given the AGE and SQFT variables, base regression is;

$$PRICE_i = \beta_1 + \beta_2\, SQFT + \beta_3 \cdot AGE + u_i$$

c) For a house away from UTOWN; it is $\hat{\beta}_2 = 76{,}12$

For a house at UTOWN; it is $\hat{\beta}_2 + \hat{\lambda} = 76{,}12 + 12{,}99 = 89{,}11$

Note that, SQFT*UTOWN variable stands for interaction effect of being in UTOWN and a unit increase in SQFT, on price of the house. Namely, (if λ is significant), contribution of being at UTOWN has an extra increase of price by 12,99\$ caused by a unit increase in SQFT.

d) (i) $H_0: \lambda = 0$
$H_A: \lambda \neq 0$
$\alpha = 0{,}05$

(ii) $t = \dfrac{\hat{\beta} - \lambda}{SE(\hat{\lambda})}$; $df \to \infty$

(iii) Reject H_0 if p-Value $< \alpha$

(iv) p-Value = 0,0001

(v) Reject H_0. Marginal effect of SQFT depends on location at $\alpha = 0{,}05$ (also at $\alpha = 0{,}001$, a very powerful reasoning)

e) $\hat{\beta}_3 = -190{,}09\$$. Yes, it is significant since this coefficient has a very low p-value = 0,0002.

g) (i) $H_0: \delta_2 \leq 0$
$H_A: \delta_2 > 0$
$\alpha = 0{,}05$

(ii) $t = \dfrac{\hat{\delta}_2 - \delta_2}{SE(\hat{\delta}_2)}$; $df \to \infty$

(iii) Reject H_0 if p-value $< \alpha$

(iv) p-Value = 2 . 0,0003 = 0,0006
since we make a one sided test.

(v) Reject H_0. Pool has a positive impact on Prices at $\alpha = 0{,}05$ (Also at $\alpha = 0{,}001$)

f) (i) $H_0: \delta_3 \leq 0$
$H_A: \delta_3 > 0$
$\alpha = 0{,}05$

(ii) $t = \dfrac{\hat{\delta}_3 - \delta_3}{SE(\hat{\delta}_3)}$; $df \to \infty$

(iii) Reject H_0 if p-Value $< \alpha$

(iv) p-Value = 2 . 0,0901 = 0,1802

(v) Do NOT Reject H_0. Fireplace has NOT a significant positive impact on Price at $\alpha = 0{,}05$

h) (i) $H_0: \delta_2 \leq \delta_3$
$H_A: \delta_2 > \delta_3$
$\alpha = 0{,}05$

(ii) $t = \dfrac{\hat{\delta}_2 - \hat{\delta}_3}{SE(\hat{\delta}_2 - \hat{\delta}_3)}$; $df \to \infty$

(iii)


Reject H_0 if $t > 1{,}645$

(iv) $Var(\hat{\delta}_2 - \hat{\delta}_3)$
$= Var(\hat{\delta}_2) + Var(\hat{\delta}_3) - 2\underbrace{Cov(\hat{\delta}_2, \hat{\delta}_3)}_{?}$

We can NOT continue this test since Cov is unknown.

i) Both δ_1 and λ are significant, location effects house prices both in intercept and in slope. They have dissimilar regressions.

j) $PRICE_t = \beta_1 + \delta_2\, POOL + \lambda\,(AGE * POOL) + \beta_3\, AGE$