ACE 564
Spring 2006

Lecture 7

Extensions of The Multiple Regression Model: Dummy Independent Variables

by
Professor Scott H. Irwin

Readings:

Griffiths, Hill and Judge. "Dummy Variables and Varying Coefficient Models" Chapter 12 in *Learning and Practicing Econometrics*

Kennedy. "Dummy Variables," Chapter 14 in *A Guide to Econometrics*

Introduction

Independent variables in regression models sometimes are "qualitative" in nature

- Male vs. female
- Wartime vs. peacetime
- Farmer vs. non-farmer
- Region 1 vs. Region 2
- Corn vs. soybeans

The qualitative nature of these variables means some type of proxy must be constructed

Definition:

A dummy variable is an artificial variable constructed to take on the value of one when the qualitative phenomenon it represents occurs, and zero otherwise

Once created, dummy variables can be used in the multiple linear regression model just like other variables

$\Rightarrow$ However, will change interpretation of model!

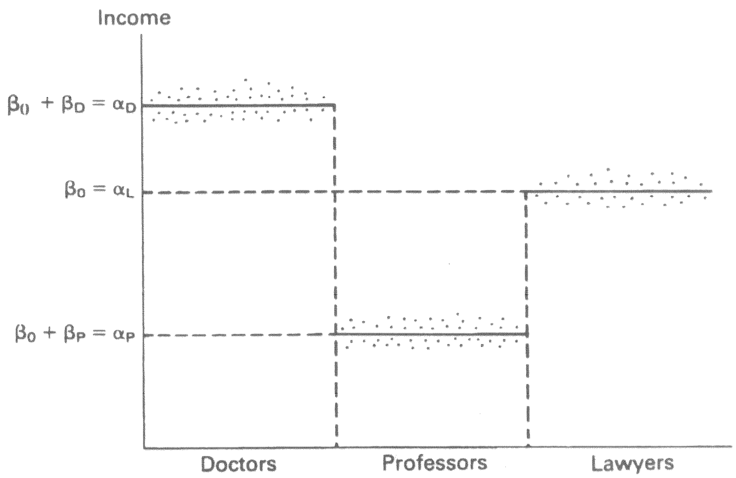


Figure 14.1 A step function example of using dummy variables

Kennedy, P. *Guide to Econometrics*, Fourth Edition. The MIT Press, Cambridge, Mass. 1998.

Dummy Variables and Income of Professionals

Data on income of professionals as shown on previous page

Assume an individual's income depends on their profession, so

$$y_t = \alpha_1 D_{1,t} + \alpha_2 D_{2,t} + \alpha_3 D_{3,t} + e_t$$

where

y_t is an individual's annual income

$D_{1,t}$ is a dummy variable taking on a value of one if a person is a doctor, zero otherwise

$D_{2,t}$ is a dummy variable taking on a value of one if a person is a professor, zero otherwise

$D_{3,t}$ is a dummy variable taking on a value of one if a person is a lawyer, zero otherwise

Now, let's consider the form of the model when we limit it to a particular profession

Doctor:

$$y_t = \alpha_1 \cdot 1 + \alpha_2 \cdot 0 + \alpha_3 \cdot 0 + e_t$$

$$y_t = \alpha_1 + e_t$$

Professor:

$$y_t = \alpha_1 \cdot 0 + \alpha_2 \cdot 1 + \alpha_3 \cdot 0 + e_t$$

$$y_t = \alpha_2 + e_t$$

Lawyer:

$$y_t = \alpha_1 \cdot 0 + \alpha_2 \cdot 0 + \alpha_3 \cdot 1 + e_t$$

$$y_t = \alpha_3 + e_t$$

Look familiar? Just the constant mean model we studied in ACE 362

Hypothetical Data Set for Professional Incomes

y_t	$D_{1,t}$	$D_{2,t}$	$D_{3,t}$
\$50,000	1	0	0
\$55,000	1	0	0
\$75,000	1	0	0
\$35,000	0	1	0
\$38,000	0	1	0
\$30,000	0	1	0
\$45,000	0	0	1
\$47,000	0	0	1
\$41,000	0	0	1

Estimate constant mean by sample average for each profession

	$\bar{x}$
Doctors	$60,000.00
Professors	$34,333.33
Lawyers	$44,333.33

Turns out we get exactly the same parameter estimates if we estimate the full model with least squares!

$$\hat{y}_t = 60{,}000 D_{1,t} + 34{,}333.33 D_{2,t} + 44{,}333.33 D_{3,t}$$

(12.71) (7.27) (9.39) (t-stat.)

Dummy variable parameter estimates in this regression show the average, or expected, income for each of the professions

Notice that an intercept in the usual sense is not included in the regression model

This is done to avoid the "dummy variable trap"

- The intercept (β_1) in a regression model is implicitly represented by a column of ones

- By definition,

$$1 = D_1 + D_2 + D_3$$

- Hence, if an intercept is included in the model an exact linear relationship exists between the dummy variables and the implicit intercept variable of the regression model

- Impossible to estimate model parameters with least squares

It is typical to include an intercept in regression models with dummy variables

- But, one of the dummies has to be dropped in order to avoid the dummy variable trap

- Changes the interpretation of the dummy parameter estimates!

Let's consider this approach by dropping the dummy variable for doctors from the model and changing notation to keep the different dummy variable models straight

$$y_t = \beta_1 + \beta_2 D_{2,t} + \beta_3 D_{3,t} + e_t$$

In this formulation:

- The income of the omitted category (doctors) is given by the intercept

- The income of the included categories (professors and lawyers) is given by the sum of the intercept and relevant slope

Doctor:

$$y_t = \beta_1 + \beta_2 \cdot 0 + \beta_3 \cdot 0 + e_t$$

$$y_t = \beta_1 + e_t$$

Professor:

$$y_t = \beta_1 + \beta_2 \cdot 1 + \beta_3 \cdot 0 + e_t$$

$$y_t = \beta_1 + \beta_2 + e_t$$

Lawyer:

$$y_t = \beta_1 + \beta_2 \cdot 0 + \beta_3 \cdot 1 + e_t$$

$$y_t = \beta_1 + \beta_3 + e_t$$

If we estimate this formulation with least squares, the results are

$$\hat{y}_t = 60{,}000 - 25{,}666.77 D_{2,t} - 15{,}666.77 D_{3,t}$$

$$(12.71) \quad (-3.84) \quad (-2.35) \quad (t\text{-stat.})$$

The estimated expected incomes for the three professions are,

Doctor:

$$\hat{y}_t = \$60{,}000$$

Professor:

$$\hat{y}_t = 60{,}000 - 25{,}666.77 = \$34{,}333.33$$

Lawyer:

$$\hat{y}_t = 60{,}000 - 15{,}666.77 = \$44{,}333.33$$

We can now see that the two formulations of the dummy variable regression model provide the same information regarding expected incomes of professionals, with proper interpretation

With dummy variables for all three professions and no intercept, slopes directly estimate expected incomes of the different professions

With one dummy variable omitted, the omitted profession becomes the benchmark to which the others are compared

More formally, the equivalence between the parameters of the two regression models is,

$$\alpha_1 = \beta_1$$

$$\alpha_2 = \beta_1 + \beta_2$$

$$\alpha_3 = \beta_1 + \beta_3$$

Most researchers prefer the intercept version, as it directly shows whether the categorization makes a difference and by how much

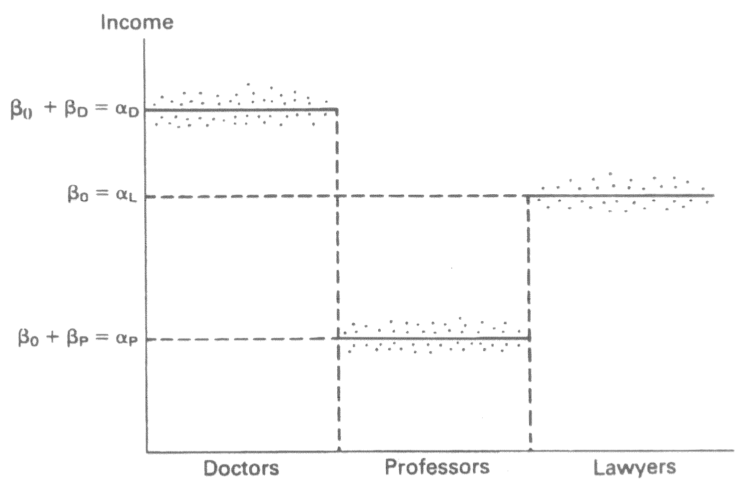


Figure 14.1 A step function example of using dummy variables

Kennedy, P. *Guide to Econometrics*, Fourth Edition. The MIT Press, Cambridge, Mass. 1998.

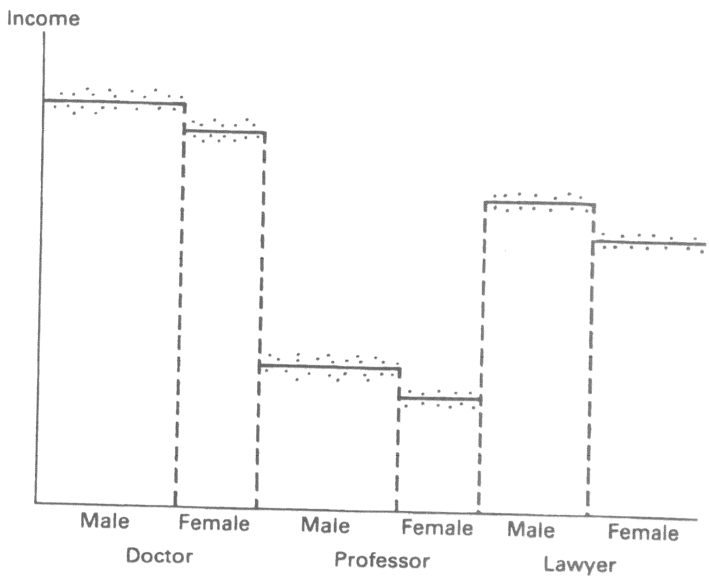


Figure 14.2 Adding gender as an additional dummy variable

Kennedy, P. *Guide to Econometrics*, Fourth Edition. The MIT Press, Cambridge, Mass. 1998.

To estimate the impact of gender on expected income, we add one more dummy variable to the regression model

$$y_t = \beta_1 + \beta_2 D_{2,t} + \beta_3 D_{3,t} + \beta_4 D_{4,t} + e_t$$

where

y_t is an individual's annual income

$D_{2,t}$ is a dummy variable taking on a value of one if a person is a professor, zero otherwise

$D_{3,t}$ is a dummy variable taking on a value of one if a person is a lawyer, zero otherwise

$D_{4,t}$ is a dummy variable taking on a value of one if a person is female, zero if a person is male

Note that only one dummy variable is added for gender

If we created one dummy for males and one for females, then we would be back in the dummy variable trap because $D_M + D_F$ equals one, the same value as the implicit intercept variable

We can now work out the model by profession and gender

Male Doctor:

$$y_t = \beta_1 + \beta_2 \cdot 0 + \beta_3 \cdot 0 + \beta_4 \cdot 0 + e_t$$

$$y_t = \beta_1 + e_t$$

Female Doctor:

$$y_t = \beta_1 + \beta_2 \cdot 0 + \beta_3 \cdot 0 + \beta_4 \cdot 1 + e_t$$

$$y_t = \beta_1 + \beta_4 + e_t$$

Male Professor:

$$y_t = \beta_1 + \beta_2 \cdot 1 + \beta_3 \cdot 0 + \beta_4 \cdot 0 + e_t$$

$$y_t = \beta_1 + \beta_2 + e_t$$

Female Professor

$$y_t = \beta_1 + \beta_2 \cdot 1 + \beta_3 \cdot 0 + \beta_4 \cdot 1 + e_t$$

$$y_t = \beta_1 + \beta_2 + \beta_4 + e_t$$

Male Lawyer:

$$y_t = \beta_1 + \beta_2 \cdot 0 + \beta_3 \cdot 1 + \beta_4 \cdot 0 + e_t$$

$$y_t = \beta_1 + \beta_3 + e_t$$

Female Lawyer:

$$y_t = \beta_1 + \beta_2 \cdot 0 + \beta_3 \cdot 1 + \beta_4 \cdot 1 + e_t$$

$$y_t = \beta_1 + \beta_3 + \beta_4 + e_t$$

Hypothetical Data Set for Professional Incomes

y_t	$D_{2,t}$	$D_{3,t}$	$D_{4,t}$
$50,000	0	0	1
$55,000	0	0	1
$75,000	0	0	0
$35,000	1	0	1
$38,000	1	0	0
$30,000	1	0	1
$45,000	0	1	0
$47,000	0	1	0
$41,000	0	1	1

If we estimate this expanded model with least squares, the results are

$$\hat{y}_t = 67,333.33 - 25,666.67 D_{2,t} - 19,333.33 D_{3,t} - 11,000.00 D_{4,t}$$
$$(16.00) \qquad (-5.57) \qquad (-4.03) \qquad (-2.76) \qquad (t\text{-stat.})$$

The estimated expected incomes for the professions and gender are,

Male Doctor:

$$\hat{y}_t = \$67,333.33$$

Female Doctor:

$$\hat{y}_t = 67,333.33 - 11,000.00 = \$56,333.33$$

Male Professor:

$$\hat{y}_t = 67,333.33 - 25,666.67 = \$41,666.67$$

Female Professor:

$$\hat{y}_t = 67,333.33 - 25,666.67 - 11,000.00 = \$30,666.67$$

Male Lawyer:

$$\hat{y}_t = 67,333.33 - 19,333.33 = \$48,000.00$$

Female Lawyer:

$$\hat{y}_t = 67,333.33 - 19,333.33 - 11,000.00 = \$37,000.00$$

Note that the regression income estimate for a professional group does not equal the corresponding sample average; this only occurs in the special case illustrated earlier in this section

Also note that the formulation in this section forces the gender income differential to be the same for the different professions

- We could specify a more flexible model that allowed the gender differential to vary by profession

- Allow "interaction" effects between profession and gender

- Need to specify separate dummy variables for each category of profession and gender (Kennedy, pp. 223-224)

Dummy Variables, Income of Professionals, and Years of Experience

The previous example is unrealistic, in that all of the independent variables are dummy variables

In most applications involving dummy variables, the model will also include other quantitative independent variables

This can be illustrated by expanding the income of professionals model to include years of professional work experience

- Gender not included to keep the model simple

The expanded model is,

$$y_t = \beta_1 + \beta_2 D_{2,t} + \beta_3 D_{3,t} + \beta_4 x_{4,t} + e_t$$

where the other variables are the same, except $x_{4,t}$ represents the number of years of professional experience

This model, in essence, expresses income as a linear function of experience, with a different intercept for each profession

More specifically, the incomes for the different professions are now given as,

Doctor:

$$y_t = \beta_1 + \beta_2 \cdot 0 + \beta_3 \cdot 0 + \beta_4 x_{4,t} + e_t$$

$$y_t = \beta_1 + \beta_4 x_{4,t} + e_t$$

Professor:

$$y_t = \beta_1 + \beta_2 \cdot 1 + \beta_3 \cdot 0 + \beta_4 x_{4,t} + e_t$$

$$y_t = \beta_1 + \beta_2 + \beta_4 x_{4,t} + e_t$$

Lawyer:

$$y_t = \beta_1 + \beta_2 \cdot 0 + \beta_3 \cdot 1 + \beta_4 x_{4,t} + e_t$$

$$y_t = \beta_1 + \beta_3 + \beta_4 x_{4,t} + e_t$$

Hypothetical Data Set for Professional Incomes

y_t	$D_{2,t}$	$D_{3,t}$	$x_{4,t}$
$50,000	0	0	3
$55,000	0	0	5
$75,000	0	0	15
$35,000	1	0	5
$38,000	1	0	20
$30,000	1	0	1
$45,000	0	1	6
$47,000	0	1	12
$41,000	0	1	2

If we estimate this formulation with least squares, the results are

$$\hat{y}_t = 53{,}764.14 - 26{,}480.04 D_2 - 14{,}853.29 D_3 + 813.37 x_{4,t}$$
$$(12.54) \qquad (-5.39) \qquad (-3.02) \qquad (2.48) \quad (t\text{-}stat.)$$

The estimated expected income functions for the three professions are,

Doctor:

$$\hat{y}_t = 53{,}764.14 + 813.37 x_{4,t}$$

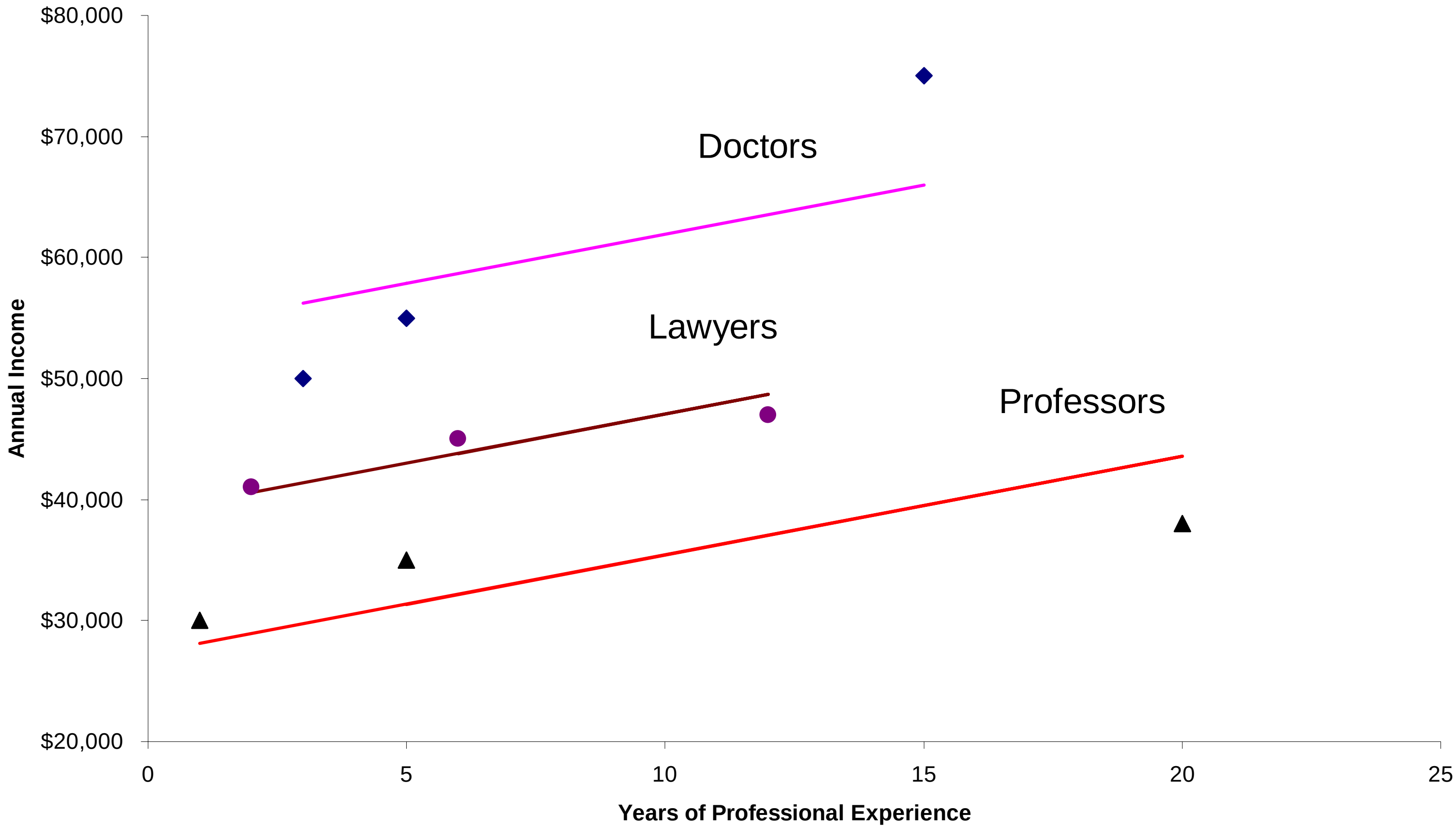
Professor:

$$\hat{y}_t = 53{,}764.14 - 26{,}480.04 + 813.37 x_{4,t} = 27{,}284.10 + 813.37 x_{4,t}$$

Lawyer:

$$\hat{y}_t = 53{,}764.14 - 14{,}853.29 + 813.37 x_{4,t} = 38{,}910.85 + 813.37 x_{4,t}$$

Estimated Regression Models for Professional Income With Intercept Dummy Variables



Testing for Individual Qualitative Effects

The point estimates for the regression model indicate that,

- Holding experience constant, doctors earn \$26,480 more per year than professors

- Holding experience constant, doctors earn \$14,853 more per year than lawyers

We would like to test whether these income differences are significantly different from a statistical standpoint

We start by re-stating the estimation results,

$$\hat{y}_t = 53{,}764.14 - 26{,}480.04 D_2 - 14{,}853.29 D_3 + 813.37 x_{4,t}$$
$$(12.54) \qquad (-5.39) \qquad (-3.02) \qquad (2.48) \qquad (t\text{-}stat.)$$

Problem I: Testing Significance of Difference in the Income of Doctors and Professors

1. Hypotheses

$$H_0 : \beta_2 = 0$$
$$H_1 : \beta_2 \neq 0$$

2. Test statistic

$$t_2 = \frac{b_2 - \beta_2^*}{se(b_2)} = \frac{b_2 - 0}{se(b_2)} = \frac{b_2}{se(b_2)} = \frac{-26,480.04}{4,914.97} = -5.39$$

3. Rejection Region

- For $\alpha = 0.05$, we want to find the critical *t*-value where $P[|t_2| \leq t_c] = 0.05$

- Since this is a two-tailed test, the rejection region is the two-tailed region determined as,

$$-t_{\alpha/2,T-K} \leq t_2 \leq t_{\alpha/2,T-K}$$
$$-t_{0.05/2,9-4} \leq t_2 \leq t_{0.05/2,9-4}$$
$$-2.571 \leq t_2 \leq 2.571$$

4. Decision

- Since $(t_2 = -5.39) < (t_c = -2.571)$ we reject the null hypothesis and conclude that the alternative hypothesis is more consistent with the sample data

- Sample evidence supports the proposition of a statistically significant difference in the annual income of doctors and professors

Problem II: Testing Significance of Difference in the Income of Doctors and Lawyers

1. Hypotheses

$$H_0 : \beta_3 = 0$$
$$H_1 : \beta_3 \neq 0$$

2. Test statistic

$$t_2 = \frac{b_3 - \beta_3^*}{se(b_3)} = \frac{b_3 - 0}{se(b_3)} = \frac{b_3}{se(b_3)} = \frac{-14{,}853.29}{4{,}914.97} = -3.02$$

3. Rejection Region

- For $\alpha = 0.05$, we want to find the critical *t*-value where $P[|t_2| \leq t_c] = 0.05$

- Since this is a two-tailed test, the rejection region is the two-tailed region determined as,

$$
\begin{aligned}
&-t_{\alpha/2,T-K} \leq t_2 \leq t_{\alpha/2,T-K} \\
&-t_{0.05/2,9-4} \leq t_2 \leq t_{0.05/2,9-4} \\
&-2.571 \leq t_2 \leq 2.571
\end{aligned}
$$

4. Decision

- Since $(t_2 = -3.02) < (t_c = -2.571)$ we reject the null hypothesis and conclude that the alternative hypothesis is more consistent with the sample data

- Sample evidence supports the proposition of a statistically significant difference in the annual income of doctors and lawyers

Testing Jointly for Qualitative Effects

Now we want to test whether the income differences are jointly significantly different

This test is just another application of the *F*-test methodology presented in Lecture 6

The regression model with a full set of intercept of dummy variables is specified as the unrestricted regression model,

$$y_t = \beta_1 + \beta_2 D_{2,t} + \beta_3 D_{3,t} + \beta_4 x_{4,t} + e_t$$

When estimated, this model will have sum of squared errors SSE_U

The null hypothesis of interest is,

$$H_0 : \beta_2 = \beta_3 = 0$$

Now, impose the null hypothesis restrictions on the unrestricted regression model to obtain the following restricted model,

$$y_t = \beta_1 + \beta_4 x_{4,t} + e_t$$

When estimated, this model will have sum of squared errors SSE_R

We can then compute the test statistic F,

$$F = \frac{(SSE_R - SSE_U)/J}{SSE_U/(T-K)}$$

Finally, compare the calculated F to the critical value from an F-distribution table, $F_\alpha(J, T-K)$

Unrestricted model estimates:

$$\hat{y}_t = 53{,}764.14 - 26{,}480.04 D_2 - 14{,}853.29 D_3 + 813.37 x_{4,t}$$

$$(12.54) \quad (-5.39) \quad (-3.02) \quad (2.48) \quad (t\text{-}stat.)$$

$$SSE_U = 180{,}366{,}932.8$$

Restricted model estimates:

$$\hat{y}_t = 45{,}757.23 + 53.65 x_{4,t}$$

$$(25.44) \quad (-5.78) \quad (t\text{-stat.})$$

$$SSE_R = 1{,}404{,}714{,}992$$

$$F = \frac{(1{,}404{,}714{,}992 - 180{,}366{,}932.8)/2}{180{,}366{,}932.8/(9-4)} = 16.97$$

Since $F = 16.97 > F_{0.05}(2,5) = 5.79$, reject null hypothesis

- Sample data are not consistent with the hypothesis that the intercept parameters are the same for all three professions

- Inappropriate to restrict the intercept to be the same for doctors, professors and lawyers

Intercept and Slope Dummy Variables, Income of Professionals and Years of Experience

We are not limited to regression models where only the intercept is allowed to shift

It is possible to allow the slope on non-dummy independent variables to shift as well

The setup for this type of regression model is,

$$y_t = \beta_1 + \beta_2 D_{2,t} + \beta_3 D_{3,t} + \beta_4 x_{4,t} + \beta_5 (D_{2,t} x_{4,t}) + \beta_6 (D_{3,t} x_{4,t}) + e_t$$

where two new variables, $(D_{2,t} x_{4,t})$ and $(D_{3,t} x_{4,t})$, are included to represent the "interaction" of profession and years of professional experience on income

$\Rightarrow$Experience may be more valuable in some professions

- β_5 is the difference between slope for doctors and professors
- β_6 is the difference between slope for doctors and lawyers

More specifically, the incomes for the different professions are now given as,

Doctor:

$$y_t = \beta_1 + \beta_2 \cdot 0 + \beta_3 \cdot 0 + \beta_4 x_{4,t} + \beta_5 (0 \cdot x_{4,t}) + \beta_6 (0 \cdot x_{4,t}) + e_t$$

$$y_t = \beta_1 + \beta_4 x_{4,t} + e_t$$

Professor:

$$y_t = \beta_1 + \beta_2 \cdot 1 + \beta_3 \cdot 0 + \beta_4 x_{4,t} + \beta_5 (1 \cdot x_{4,t}) + \beta_6 (0 \cdot x_{4,t}) + e_t$$

$$y_t = \beta_1 + \beta_2 + (\beta_4 + \beta_5) x_{4,t} + e_t$$

Lawyer:

$$y_t = \beta_1 + \beta_2 \cdot 0 + \beta_3 \cdot 1 + \beta_4 x_{4,t} + \beta_5 (0 \cdot x_{4,t}) + \beta_6 (1 \cdot x_{4,t}) + e_t$$

$$y_t = \beta_1 + \beta_3 + (\beta_4 + \beta_6) x_{4,t} + e_t$$

Hypothetical Data Set for Professional Incomes

y_t	$D_{2,t}$	$D_{3,t}$	$x_{4,t}$	$D_{2,t}x_{4,t}$	$D_{3,t}x_{4,t}$
$50,000	0	0	3	0	0
$55,000	0	0	5	0	0
$75,000	0	0	15	0	0
$35,000	1	0	5	5	0
$38,000	1	0	20	20	0
$30,000	1	0	1	1	0
$45,000	0	1	6	0	6
$47,000	0	1	12	0	12
$41,000	0	1	2	0	2

If we estimate this formulation with least squares, the results are

$$\hat{y}_t = 44,233.87 - 13,024.57D_2 - 3,760.10D_3 +$$
$$\quad (25.44) \qquad (-5.78) \qquad (-1.47)$$
$$2056.45x_{4,t} - 1,695.99(D_{2,t}x_{4,t}) - 1,477.50(D_{3,t}x_{4,t})$$
$$\quad (10.99) \qquad (-7.63) \qquad (-4.87) \qquad (t\text{-stat.})$$

The estimated expected income functions for the three professions are,

Doctor:

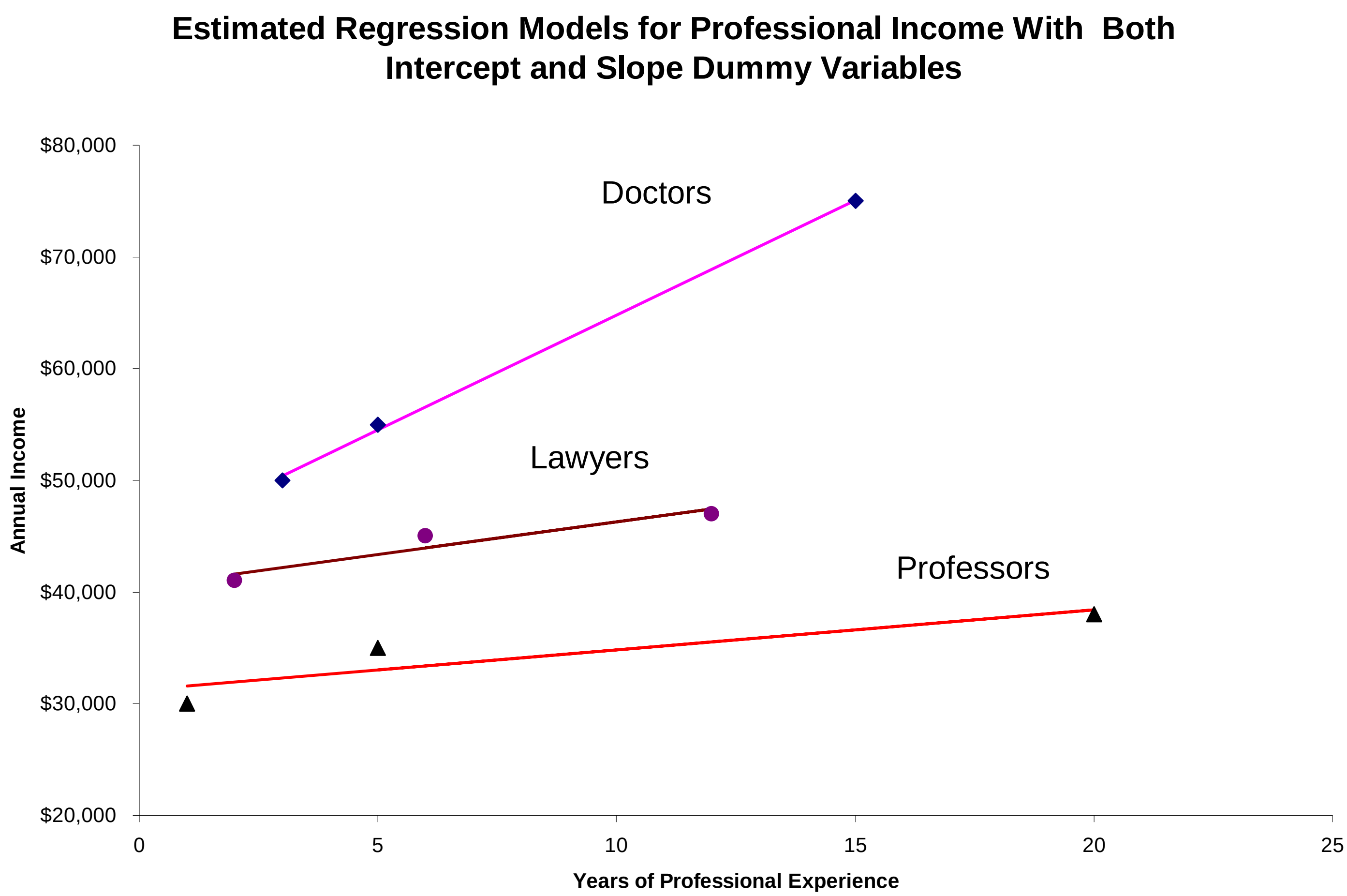
$$\hat{y}_t = 44,233.87 + 2,056.45x_{4,t}$$

Professor:

$$\begin{aligned}\hat{y}_t &= 44,233.87 - 13,024.57 + (2,056.45 - 1,695.99)x_{4,t} \\ &= 31,209.30 + 360.46x_{4,t}\end{aligned}$$

Lawyer:

$$\begin{aligned}\hat{y}_t &= 44,233.87 - 3,760.10 + (2,056.45 - 1,477.50)x_{4,t} \\ &= 40,473.77 + 578.95x_{4,t}\end{aligned}$$


Estimated Regression Models for Professional Income With Both Intercept and Slope Dummy Variables
Doctors
Lawyers
Professors
$80,000
$70,000
$60,000
$50,000
$40,000
$30,000
$20,000
Annual Income
0
5
10
15
20
25
Years of Professional Experience

The Chow Test

We would like to formally test whether,

> Intercepts and slopes are different for each of the three groups of professionals

vs.

> Intercepts and slopes are the same for each of the three groups of professionals

This test is just another application of the *F*-test methodology presented in Lecture 6

- Named after econometrician Gregory Chow, who first proposed the test

- Also, known as a test of the equivalence of regressions, a test of pooling data or at test of structural change

The regression model with a full set of intercept and slope dummy variables is specified as the unrestricted regression model,

$$y_t = \beta_1 + \beta_2 D_{2,t} + \beta_3 D_{3,t} + \beta_4 x_{4,t} + \beta_5 (D_{2,t} x_{4,t}) + \beta_6 (D_{3,t} x_{4,t}) + e_t$$

When estimated, this model will have sum of squared errors SSE_U

The null hypothesis of interest is,

$$H_0 : \beta_2 = \beta_3 = \beta_5 = \beta_6 = 0$$

Now, impose the null hypothesis restrictions on the unrestricted regression model to obtain the following restricted model,

$$y_t = \beta_1 + \beta_4 x_{4,t} + e_t$$

When estimated, this model will have sum of squared errors SSE_R

We can then compute the test statistic F,

$$F = \frac{(SSE_R - SSE_U)/J}{SSE_U/(T-K)}$$

Finally, compare the calculated F to the critical value from an F-distribution table, $F_\alpha(J, T-K)$

Unrestricted model estimates:

$$\hat{y}_t = 44,233.87 - 13,024.57 D_2 - 3,760.10 D_3 +$$
$$(25.44) \quad (-5.78) \quad (-1.47)$$
$$2056.45 x_{4,t} - 1,695.99 (D_{2,t} x_{4,t}) - 1,477.50 (D_{3,t} x_{4,t})$$
$$(10.99) \quad (-7.63) \quad (-4.87) \quad (t\text{-stat.})$$
$$SSE_U = 8,680,459.589$$

Restricted model estimates:

$$\hat{y}_t = 45,757.23 + 53.65 x_{4,t}$$
$$(25.44) \quad (-5.78) \quad (t\text{-stat.})$$
$$SSE_R = 1,404,714,992$$

$$F = \frac{(1,404,714,992 - 8,680,459.589)/4}{8,680,459.589/(9-6)} = 120.62$$

Since $F = 120.62 > F_{0.05}(4,3) = 9.13$, reject null hypothesis

- Sample data are not consistent with the hypothesis that the intercept and slope parameters are the same for all three professions

- Inappropriate to pool the sample data for doctors, professors and lawyers and estimate one regression model

While it is clearly inappropriate to estimate one regression for all three professions, we do not yet know whether we can restrict the intercepts or slopes (but not both) to be the same for the different professions

These tests are not Chow tests for pooling but are so close in procedure that this is a good place to go over such testing

Testing for Equivalence of Slopes

Once again, the regression model with a full set of intercept and slope dummy variables is specified as the unrestricted regression model,

$$y_t = \beta_1 + \beta_2 D_{2,t} + \beta_3 D_{3,t} + \beta_4 x_{4,t} + \beta_5 (D_{2,t} x_{4,t}) + \beta_6 (D_{3,t} x_{4,t}) + e_t$$

When estimated, this model will have sum of squared errors SSE_U

The null hypothesis of interest is,

$$H_0 : \beta_5 = \beta_6 = 0$$

Now, impose the null hypothesis restrictions on the unrestricted regression model to obtain the following restricted model,

$$y_t = \beta_1 + \beta_2 D_{2,t} + \beta_3 D_{3,t} + \beta_4 x_{4,t} + e_t$$

When estimated, this model will have sum of squared errors SSE_R

We can then compute the test statistic F,

$$F = \frac{(SSE_R - SSE_U)/J}{SSE_U/(T-K)}$$

Finally, compare the calculated F to the critical value from an F-distribution table, $F_\alpha(J, T-K)$

Unrestricted model estimates:

$$\hat{y}_t = 44{,}233.87 - 13{,}024.57D_2 - 3{,}760.10D_3 +$$
$$\quad (25.44) \qquad (-5.78) \qquad (-1.47)$$
$$2056.45x_{4,t} - 1{,}695.99(D_{2,t}x_{4,t}) - 1{,}477.50(D_{3,t}x_{4,t})$$
$$(10.99) \qquad (-7.63) \qquad (-4.87) \qquad (t\text{-stat.})$$
$$SSE_U = 8{,}680{,}459.589$$

Restricted model estimates:

$$\hat{y}_t = 53{,}764.14 - 26{,}480.04D_2 - 14{,}853.29D_3 + 813.37x_{4,t}$$
$$(12.54) \qquad (-5.39) \qquad (-3.02) \qquad (2.48) \qquad (t\text{ - }stat.)$$
$$SSE_R = 180{,}366{,}932.8$$

$$F = \frac{(180{,}366{,}932.8 - 8{,}680{,}459.589)/2}{8{,}680{,}459.589/(9-6)} = 29.67$$

Since $F = 29.67 > F_{0.05}(2,3) = 9.55$, reject null hypothesis

- Sample data are not consistent with the hypothesis that the slope parameters are the same for all three professions

- Inappropriate to restrict the slopes to be the same for doctors, professors and lawyers

Testing for Equivalence of Intercepts

Once again, the regression model with a full set of intercept and slope dummy variables is specified as the unrestricted regression model,

$$y_t = \beta_1 + \beta_2 D_{2,t} + \beta_3 D_{3,t} + \beta_4 x_{4,t} + \beta_5 (D_{2,t} x_{4,t}) + \beta_6 (D_{3,t} x_{4,t}) + e_t$$

When estimated, this model will have sum of squared errors SSE_U

The null hypothesis of interest is,

$$H_0 : \beta_2 = \beta_3 = 0$$

Now, impose the null hypothesis restrictions on the unrestricted regression model to obtain the following restricted model,

$$y_t = \beta_1 + \beta_4 x_{4,t} + \beta_5 (D_{2,t} x_{4,t}) + \beta_6 (D_{3,t} x_{4,t}) + e_t$$

When estimated, this model will have sum of squared errors SSE_R

We can then compute the test statistic F,

$$F = \frac{(SSE_R - SSE_U)/J}{SSE_U\ /(T-K)}$$

Finally, compare the calculated F to the critical value from an F-distribution table, $F_\alpha(J, T-K)$

Unrestricted model estimates:

$$\begin{aligned}\hat{y}_t = &\ 44,233.87 - 13,024.57 D_2 - 3,760.10 D_3 + \\ &\ (25.44) \qquad (-5.78) \qquad (-1.47) \\ &\ 2056.45 x_{4,t} - 1,695.99(D_{2,t} x_{4,t}) - 1,477.50(D_{3,t} x_{4,t}) \\ &\ (10.99) \qquad (-7.63) \qquad (-4.87) \qquad (t\text{-stat.}) \\ &\ SSE_U = 8,680,459.589\end{aligned}$$

Restricted model estimates:

$$\begin{aligned}\hat{y}_t = &\ 37504.61 + 2,654.03 x_{4,t} - 2,677.79(D_{2,t} x_{4,t}) - 1,752.36(D_{3,t} x_{4,t}) \\ &\ (13.96) \qquad (6.95) \qquad (-6.94) \qquad (\text{-}3.76) \qquad (t\text{ - }stat.) \\ SSE_R = &\ 115,326,714.6\end{aligned}$$

$$F = \frac{(115,326,714.6 - 8,680,459.589)/2}{8,680,459.589/(9-6)} = 18.43$$

Since $F = 18.43 > F_{0.05}(2,3) = 9.55$, reject null hypothesis

- Sample data are not consistent with the hypothesis that the intercept parameters are the same for all three professions

- Inappropriate to restrict the intercept to be the same for doctors, professors and lawyers

It is interesting to note that hypothesis testing suggests the unrestricted model, with a full set of intercept and slope "shifters," is consistent with the sample data

With one exception, this is the same as saying that three separate regressions should be estimated for the different professions

- Unrestricted model assumes variance of regression is the same for all three professions

- Three separate regressions allows the variance of regression to differ across the professions

Structural Change and the Consumption-Income Relationship in the US

The income of professionals example is used to examine cross-sectional qualitative effects

It is useful to present an example of time-series qualitative effects

The Problem: Estimating the relationship between consumption and income in the US over 1929-1970

- Time period spans the Great Depression and several wars
- Consumption was particularly restricted relative to income during WWII
- May not be appropriate to assume the same relationship over the entire sample period

Table 12.1 U.S. Real Per Capita Income and Consumption

Year	C	Y	Year	C	Y
1929	1145	1236	1950	1520	1646
1930	1059	1128	1951	1509	1657
1931	1016	1077	1952	1525	1678
1932	919	9210	1953	1572	1726
1933	897	8930	1954	1575	1714
1934	934	9520	1955	1659	1795
1935	985	1035	1956	1673	1839
1936	1080	1158	1957	1683	1844
1937	1110	1187	1958	1666	1831
1938	1097	1105	1959	1735	1881
1939	1131	1190	1960	1749	1883
1940	1178	1259	1961	1755	1909
1941	1240	1427	1962	1813	1968
1942	1197	1582	1963	1865	2013
1943	1213	1629	1964	1945	2123
1944	1238	1673	1965	2044	2235
1945	1308	1642	1966	2123	2331
1946	1439	1606	1967	2160	2398
1947	1431	1513	1968	2248	2480
1948	1438	1567	1969	2301	2517
1949	1451	1547	1970	2323	2579

Source: Economic Report of the President, 1972.

Griffiths, William E., R. Carter Hill, George G. Judge. Learning and Practicing Econometrics. John Wiley & Sons, Inc. New York. 1993.

Begin with the following statistical model,

$$y_t = \beta_1 + \beta_3 x_t + e_t$$

where

y_t is real per capita consumption in the US

x_t is real per capita income in the US

Next, define a dummy variable to reflect the period in WWII when consumption was restricted through war-time rationing

$$D_t = \begin{cases} 1 \text{ if } t = 1941,\ldots,1946 \\ 0 \text{ otherwise} \end{cases}$$

Now, add the dummy variable to the statistical model,

$$y_t = \beta_1 + \beta_2 D_t + \beta_3 x_t + e_t$$

The model for different time periods is,

1929-1940, 1947-1970:

$$y_t = \beta_1 + \beta_2 \cdot 0 + \beta_3 x_t + e_t$$

$$y_t = \beta_1 + \beta_3 x_t + e_t$$

1941-1946:

$$y_t = \beta_1 + \beta_2 \cdot 1 + \beta_3 x_t + e_t$$

$$y_t = \beta_1 + \beta_2 + \beta_3 x_t + e_t$$

Estimation Results:

$$\hat{y}_t = 101.36 - 204.95 D_t + 0.86 x_t$$
$$(3.98) \quad (-10.91) \quad (58.73) \quad (t\text{-}stat.)$$

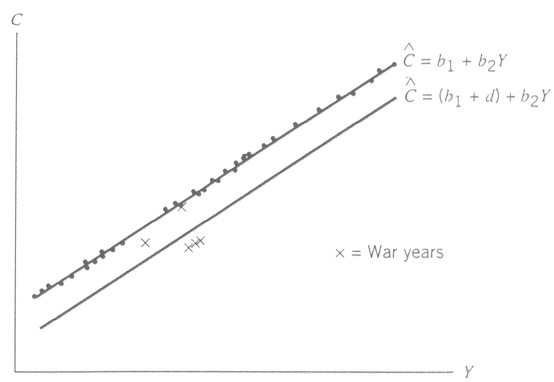


FIGURE 9.2 The estimated consumption function for the war and nonwar years

Hill, C., W. Griffiths, and G. Judge. *Undergraduate Econometrics.* John Wiley & Sons, Inc., New York, NY 1997.