Week 2: Multiple Regression, Training Exercises

Coursera/Erasumus U., Econometric Methods and Applications

Anthony Nguyen

Training Exercise 2.2

Questions

In the wage database, education is measured in terms of a single variable 'Educ' with values 1, 2, 3, and 4. The multiple regression model (with Educ as 4-th explanatory factor) assumes a constant marginal effect:

$$\frac{\partial log(Wage)}{\partial Educ} = \beta_4$$

This means that increasing education by one level always leads to the same relative wage increase. This effect may, however, depend on the education level, for example, if the effect is smaller for a shift from eduction level 1 to 2 as compared to a shift from 3 to 4.

(a) The wage equation presented at the start of Lecture 2.2 contains four explanatory factors (apart from the constant term). Formulate the null hypothesis that none of these four factors has effect on wage in the form $R\beta = r$, that is, determine R and r.

(b) Extend the wage equation presented at the start of Lecture 2.2 by allowing for education effects that depend on the education level. *Hint: Use dummy variables for education levels 2, 3, and 4.*

(c) The model of part (b) is more general than the original wage equation. The original model can be obtained from the model in part (b) by imposing linear restrictions of the type $R\beta = r$. Derive the number of restrictions (g) and determine R and r.

Answers

(a) The wage equation presented at the start of Lecture 2.2 contains four explanatory factors (apart from the constant term). Formulate the null hypothesis that none of these four factors has effect on wage in the form $\boldsymbol{R\beta} = \boldsymbol{r}$, that is, determine $\boldsymbol{R}$ and $\boldsymbol{r}$.

Our wage equation takes the form:

$$log(Wage_i) = \beta_1 + \beta_2 Female_i + \beta_3 Age_i + \beta_4 Educ_i + \beta_5 Parttime_i + \varepsilon_i$$

The null hypothesis that none of these factors has an effect on wage is therefore equivalent to:

$\beta_2 = 0$, $\beta_3 = 0$, $\beta_4 = 0$, $\beta_5 = 0$

$$\underset{R}{\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}} \underset{\beta}{\begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{pmatrix}} = \underset{r}{\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}}$$

(b) Extend the wage equation presented at the start of Lecture 2.2 by allowing for education effects that depend on the education level. *Hint: Use dummy variables for education levels 2, 3, and 4.*

As our column 'education' is expressed in four levels, or factors, we need to convert it into dummy variables to incorporate this information into our model.

Each dummy variable would equal 1 for the level concerned, and 0 otherwise. For example:

$$EduD2_i = \begin{cases} 1 & \text{if } Educ_i = 2 \text{ (level 2)}, \\ 0 & \text{otherwise (levels 1, 3, 4)}. \end{cases}$$

Although we create four dummy variables ($EduD1$, $EduD2$, $EduD3$, $EduD4$), we need to leave out one of them from our model to avoid the dummy variable/multiple colinearity trap.

Thus, extending our original wage equation to allow for education effects that depend on specific education levels gives us:

$$log(Wage_i) = \gamma_1 + \gamma_2 Female_i + \gamma_3 Age_i + \gamma_4 EduD2_i + \gamma_5 EduD3_i + \gamma_6 EduD4_i + \gamma_7 Parttime_i + \varepsilon_i$$

Written this way, the effect of moving from one specific level of education to another level on log wage is:

level 1 to level 2: γ_4
level 1 to level 3: γ_5
level 1 to level 4: γ_6

level 2 to level 3: $\gamma_5 - \gamma_4$
level 2 to level 4: $\gamma_6 - \gamma_4$
level 3 to level 4: $\gamma_6 - \gamma_5$

(c) The model of part (b) is more general than the original wage equation. The original model can be obtained from the model in part (b) by imposing linear restrictions of the type $\boldsymbol{R\beta} = \boldsymbol{r}$. Derive the number of restrictions (g) and determine $\boldsymbol{R}$ and $\boldsymbol{r}$.

Re-writing our second regression equation (using γ) in terms of the original equation (β), gives us:

level 1 to level 2: β_4
level 1 to level 3: $2\beta_4$
level 1 to level 4: $3\beta_4$

or

level 1 to level 2: $\beta_4 = \gamma_4$
level 2 to level 3: $\beta_4 = \gamma_5 - \gamma_4$
level 3 to level 4: $\beta_4 = \gamma_6 - \gamma_5$

$\gamma_5 - \gamma_4 = \gamma_4$ $\gamma_5 = 2\gamma_4$

$\gamma_6 - \gamma_5 = \gamma_4$
$\gamma_6 = \gamma_4 + \gamma_5$
$\gamma_6 = 3\gamma_4$

Thus, there are two restrictions, where:

$\gamma_6 = 3\gamma_4$
$\gamma_5 = 2\gamma_4$

Leaving R and r to take the form:

$$\underbrace{\begin{pmatrix} 0 & 0 & 0 - 2 & 1 & 0 & 0 \\ 0 & 0 & 0 - 3 & 0 & 1 & 0 \end{pmatrix}}_{R} \gamma = \underbrace{\begin{pmatrix} 0 \\ 0 \end{pmatrix}}_{r}$$