ECON 3740: INTRODUCTION TO ECONOMETRICS

INSTRUCTOR: CHAOYI CHEN
Department of Economics and Finance, University of Guelph

Lecture 18

Lecture outline

Last lecture, we studied a single dummy variable. Today, we will

- Use a dummy as explanatory variable in equation for $log(y)$
- Study dummy variables for multiple categories
  - Incorporating Ordinal Information by using dummy variables (2 specifications)
  - Add Interactions among Dummy Variables
  - Allow different slopes

MLR, Further Issue: Using Dummy Explanatory Variables in Equations for $log(y)$

- Recall the example of the housing price regression and consider the following fitted regression line

$$\widehat{log(price)} = \underset{(0.65)}{-1.35} \; \underset{(0.038)}{+0.168\log(\text{lotsize})} \; \underset{(0.093)}{+0.707\log(\text{sqrft})} \; \underset{(0.029)}{+0.027\text{bdrms}} \; \underset{(0.045)}{+0.054\text{colonial}}$$

where $n = 88$, $R^2 = 0.649$, and

$$colonial = \begin{cases} 1, \text{the house is the colonial style} \\ 0, \text{not the colonial style} \end{cases}$$

- Now,

$$\frac{\partial log(price)}{\partial colonial} = \frac{\partial price/price}{\partial colonial} = 5.4\%$$

- As the dummy for colonial style changes from 0 to 1, the house price increases by 5.4 percentage points.

MLR, Further Issue: Using Dummy Variables for Multiple Categories

- 1. Define membership in **each** category by a dummy variable
- 2. Leave out one category (which becomes the base category)
- An example (Log Hourly Wage Equation). Suppose one would like to estimate a model that allows for wage differences among four groups: married men, married women, single men, and single women.

$$\widehat{log(wage)} = \underset{(0.1)}{-0.321} \quad \underset{(0.055)}{+0.213\text{marrmale}} \quad \underset{(0.0058)}{-0.198\text{marrfem}}$$
$$\underset{(0.056)}{-0.11\text{singfem}} \quad \underset{(0.007)}{+0.079\text{educ}} \quad \underset{(0.005)}{+0.027\text{exper}}$$
$$\underset{(0.00011)}{-0.00054 exper^2} \quad \underset{(0.007)}{+0.029\text{tenure}} \quad \underset{(0.00023)}{-0.00053 tenure^2}$$

where $n = 526$ and $R^2 = 0.461$

- Holding other things fixed, married women earn 19.8% less than single men (= the base category); similarly, married men earn 21.3% more and single women earn 11.0% (¡ 19.8%) less than single men.

MLR, Further Issue: Incorporating Ordinal Information by Using Dummy Variables

- An example (City Credit Ratings and Municipal Bond Interest Rates): We can propose two specifications of the regression model.
- The first specification is

$$MBR = \beta_0 + \beta_1 CR + other\ \ factors$$

where

$$MBR = \text{municipal bond interest rate}$$
$$CR = \text{credit rating from 0-4 (0 = worst, 4 = best)}$$

- This specification would probably not be appropriate as the credit rating only contains ordinal information.
- A better way to incorporate this information is to define dummies

$$MBR = \beta_0 + \delta_1 CR_1 + \delta_2 CR_2 + \delta_3 CR_3 + \delta_4 CR_4 + other\ \ factors$$

where $CR_1, ..., CR_4$ are dummies indicating whether the particular rating applies, e.g., $CR_1 = 1$ if $CR = 1$ and CR1 $= 0$ otherwise.
- All effects are measured in comparison to the worst rating (= base category).

MLR, Further Issue: Difference Between These Two Specifications

- Specification 1: $MBR = \beta_0 + \beta_1 CR + other\ \ factors$

$$
\begin{aligned}
CR = 0 &\Longrightarrow MBR = \beta_0 \\
CR = 1 &\Longrightarrow MBR = \beta_0 + \beta_1 \\
CR = 2 &\Longrightarrow MBR = \beta_0 + 2 * \beta_1 \\
CR = 3 &\Longrightarrow MBR = \beta_0 + 3 * \beta_1 \\
CR = 4 &\Longrightarrow MBR = \beta_0 + 4 * \beta_1
\end{aligned}
$$

where the increase in MBR for each rating improvement is the same, which equals β_1.

MLR, Further Issue: Difference Between These Two Specifications Continue

- Specification 2:
  $MBR = \beta_0 + \delta_1 CR_1 + \delta_2 CR_2 + \delta_3 CR_3 + \delta_4 CR_4 + other\ \ factors$

$$
\begin{aligned}
CR_1 = CR_2 = CR_3 = CR_4 = 0 \Longrightarrow MBR &= \beta_0 \\
CR_1 = 1 \Longrightarrow MBR &= \beta_0 + \delta_1 \\
CR_2 = 1 \Longrightarrow MBR &= \beta_0 + \delta_2 \\
CR_3 = 1 \Longrightarrow MBR &= \beta_0 + \delta_3 \\
CR_4 = 1 \Longrightarrow MBR &= \beta_0 + \delta_4
\end{aligned}
$$

where the increase in MBR for each rating improvement can be different due to the arbitrariness of $\delta_1, , \delta_4$.

MLR, Further Issue: Interactions among Dummy Variables

- Reconsider the female and marital status effect on $log(wage)$ by adding the $female * married$ interaction term,

$$
\begin{array}{llll}
\widehat{log(wage)} = & \underset{(0.1)}{-0.321} & \underset{(0.056)}{-0.11 female} & \underset{(0.55)}{+0.213 married} \\
 & & \underset{(0.072)}{-0.301 female * married} & + \ldots.
\end{array}
$$

- *marrmale*: setting $married = 1$ and $female = 0$, we get $\widehat{\delta}_2 = 0.213$ same as before (page 4)
- *marrfem*: setting $married = 1$ and $female = 1$, we get $\widehat{\delta}_1 + \widehat{\delta}_2 + \widehat{\delta}_3 = -0.198$ same as before (page 4)
- *singfem*: setting $married = 0$ and $female = 1$, we get $\widehat{\delta}_1 = -0.11$ same as before (page 4)
- So these two specifications are equivalent: four categories are generated.

MLR, Further Issue: Interactions among Dummy Variables Continue

- Now, what is the meaning of the coefficient of the interations term, *female* $*$ *married*?
- Recall the regression model is

$$log(wage) = \delta_0 + \delta_1 female + \delta_2 married + \delta_3 female * married + \mu$$

- Hence,

$$E[log(wage)|\text{female=1,married=1}] = \delta_0 + \delta_1 + \delta_2 + \delta_3 \quad (1)$$
$$E[log(wage)|\text{female=1, married=0}] = \delta_0 + \delta_1 \quad (2)$$
$$E[log(wage)|\text{female=0,married=1}] = \delta_0 + \delta_2 \quad (3)$$
$$E[log(wage)|\text{female=0,married=0}] = \delta_0 \quad (4)$$

- Therefore,

$$\delta_3 = (equation(1) - equation(2)) - (equation(3) - equation(4))$$

That is δ_3 measures the difference between female and male in the difference between married and un-married (a "difference in difference" effect).

MLR, Further Issue: Allowing for Different Slopes

- Consider the model

$$log(wage) = \beta_0 + \delta_0 female + \beta_1 educ + \delta_1 female * educ + \mu$$

where

$$\beta_0 = \text{intercept of male, } \beta_1 = \text{slope of men,}$$
$$\beta_0 + \delta_0 = \text{intercept of female, } \beta_1 + \delta_1 = \text{slope of female}$$

- Interacting both the intercept and the slope with the female dummy enables one to model completely independent wage equations for men and women.
- Therefore, some interesting Hypotheses come up

$$H_0 : \delta_1 = 0$$

Under H_0, the return to education is the same for men and women.

$$H_0 : \delta_0 = \delta_1 = 0$$

Under H_0, the whole wage equation is the same for men and women.

MLR, Further Issue: Allowing for Different Slopes, Graphical Illustration

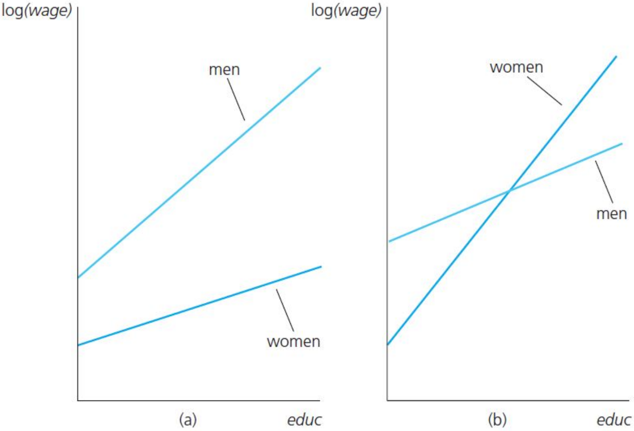

Figure: (a) $\delta_0 < 0$, $\delta_1 < 0$; (b) $\delta_0 < 0$, $\delta_1 > 0$

MLR, Further Issue: Allowing for Different Slopes, An Example

- Consider following fitted regression line

$$\widehat{log(wage)} = \underset{(0.119)}{0.389} \; \underset{(0.168)}{-0.227 female} \; \underset{(0.008)}{+0.082 educ}$$
$$\underset{(0.0131)}{-0.0056 female * educ} \; \underset{(0.005)}{+0.029 exper} \; \underset{(0.00011)}{-0.00058 exper^2}$$
$$\underset{(0.007)}{+0.032 tenure} \; \underset{(0.00024)}{-0.00059 tenure^2}$$

where $n = 526$ and $R^2 = 0.441$

- Consider the null $H_0 : \beta_{female*educ} = 0$. $|t_{female*educ}| = |\frac{-0.0056}{0.0131}| = 0.43 < 1.96$. Hence, no evidence against hypothesis that the return to education is the same for men and women.
- Consider the null $H_0 : \beta_{female} = 0$. $|t_{female}| = |\frac{-0.227}{0.168}| = 1.35 < 1.96$. Does this mean that there is no significant evidence of lower pay for women at the same levels of *educ*, *exper*, *andtenure*? **No**, this is only the effect for $educ = 0$ because

$$\frac{\partial log(wage)}{\partial female} = -0.227 - 0.0056 educ \quad (5)$$