ECO 372: Introduction to Econometrics

Spring 2025

Lecture 10: Interactions in Regressors & Introduction of Non-Linearity in Model Specifications

Sakib Bin Amin, Ph.D.

Associate Professor in Economics
Director, Accreditation Project Team (APT)

Outline

Our objectives for this lecture will be to learn:

- Understanding Interaction Effects
- Modelling Interaction Effect: Dummy and Continuous Variable
- Statistical Inference of Dummy and Continuous Variable Interaction
- Modelling Interaction Effect: Dummy Variables
- Statistical Inference from of Dummy Variable Interaction
- Non-Linearity in Causal Relationships
- A General Approach to Modeling Nonlinearities
- Visual Representations of Non-Linearity In Models
- Non-Linear Model Specifications
- Example of Non-Linear Model Estimation

What is Interaction Effect?

- Recall dummy variable modelling concepts from our last discussion.
- For instance, we said D_i is a dummy that captures wage difference of black and white employees. The coefficient of D_i remains same in the target group (black).
- However, what if there are some other factors affecting the target group, leading to some within group differences?
- Among others, the effect of education sometimes affect the expected wage within sub-groups under broad category. For example, educated black employees may earn more than less or uneducated black employees.
- When effects differ across groups or categories due to some influence, in general, we call this <u>interaction or moderation</u>.

Modelling Interaction Effect: Dummy and Continuous Variable

- Let us consider the following example:

$$Wage_i = \beta_0 + \beta_1 Edu_i + \beta_2 Black_i + u_i$$

- Now, we extend it with an interaction term:

$$Wage_i = \beta_0 + \beta_1 Edu_i + \beta_2 Black_i + \beta_3 Black_i * Edu + u_i$$

- In the above setting, β_3 captures the effect of education on the relationship of race and wage.
- β_3 is also called the <u>differential slope coefficient or slope drifter</u>. It tells us much the slope coefficient differs between two categories due to higher education.
- β_2, on the other hand, as previously refer as <u>differential intercep</u>t since it categorises sample.
- An example will make things much easier for us.

Modelling Interaction Effect: Dummy and Continuous Variable

Respondent ID	Education (Years of Education)	Race (Black=1, White=0)	Edu*Black
1	12	1	12
2	10	0	0
3	8	0	0
4	7	0	0
5	12	0	0
6	12	1	12
7	12	0	0
8	12	0	0
9	12	0	0
10	8	0	0
11	12	1	12
12	5	1	5
13	5	1	5
14	12	1	12
15	12	1	12
16	12	1	12
17	16	1	16
18	13	1	13
19	13	0	0
20	16	0	0

Modelling Interaction Effect: Dummy and Continuous Variable

```
Linear regression                               Number of obs     =     28,532
                                                F(3, 28528)       =    2076.79
                                                Prob > F          =     0.0000
                                                R-squared         =     0.1973
                                                Root MSE          =     .42838
```

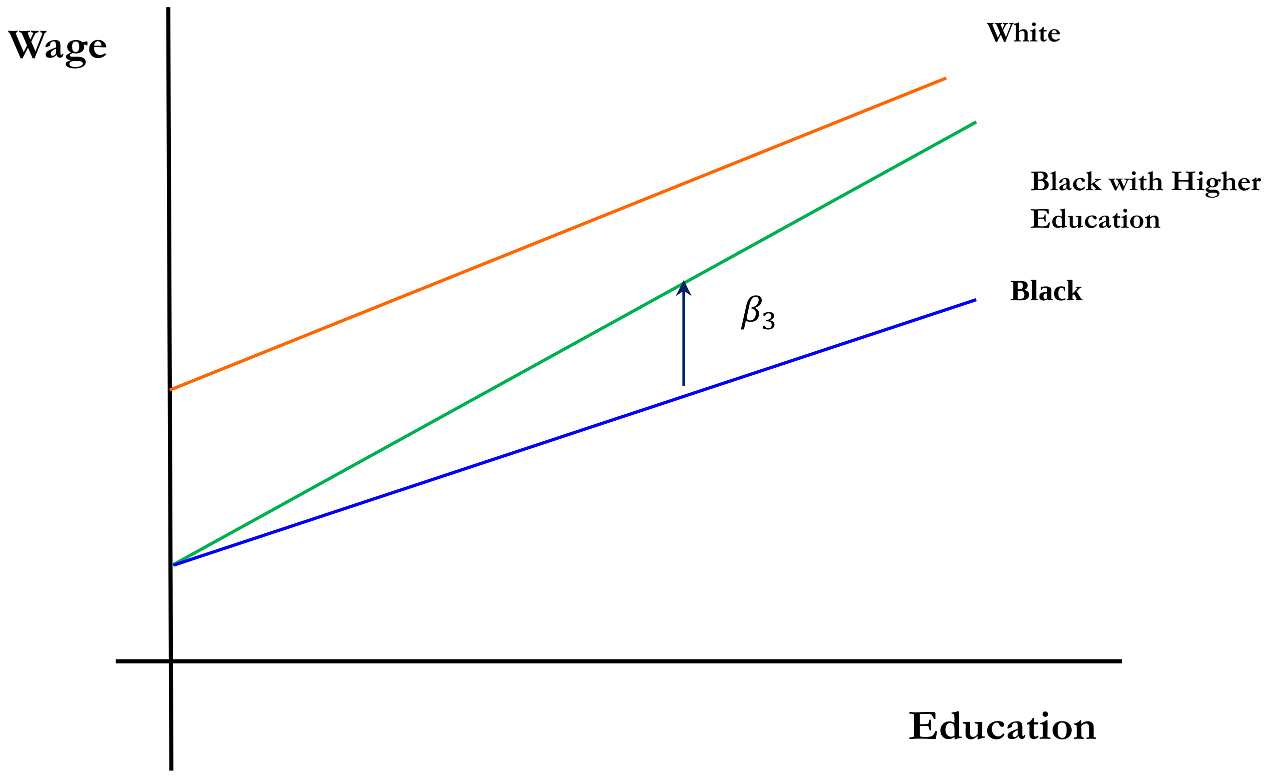
ln_wage	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
grade	.0867331	.0013898	62.41	0.000	.084009	.0894572
black	-.1294132	.0316448	-4.09	0.000	-.1914385	-.0673878
grade_black	.0050201	.0025643	1.96	0.050	-5.96e-06	.0100462
_cons	.6075497	.0177831	34.16	0.000	.572694	.6424053

STATA Commands:

webuse regsmpl
**generating interaction variable of black and edu
g grade_black=grade*black
reg ln_wage grade black grade_black, vce(r)

Modelling Interaction Effect: Dummy and Continuous Variable

- If we want to plot the model, we will get something like the figure below:



Statistical Inference from the Example

- We can see that grade or the level of education has a positive impact on wage. As education level increases by 1 unit, wage increases by around 8.67% (since wage in natural log), considering all are constant.

- It is clear that when the respondent is black, his(her) income is lower than the counterpart (i.e., white). On average, black employees earn 12.94% less than white employees, considering all are constant.

- Interestingly, higher education tends to improve wage of the Black employees by 0.5%, considering all are constant.
  - Or, earn 12.44% less than the White employees.

- This indicates that education moderates or pushes up the wage for the black educated employees, leading to decline in wage differentials.

Modelling Interaction Effect: Dummy Variables

- Similar to previous example, we can analyse the interaction of 2 or multiple dummy variables.

- Let us consider the following specification:

$$Wage_i = \beta_0 + \beta_1 Black_i + \beta_2 SMSA_i + \beta_3 SMSA_i * Black_i + \beta_4 Edu_i + u_i$$

- The specification is able to capture the wage differential of black employees who do not live in metropolitan areas compared to black employees living in metropolitans areas.

STATA Commands:

```
webuse regsmpl
**generating interaction variable of black and edu
g smsa_black= not_smsa*black
reg ln_wage black not_smsa smsa_black grade, vce(r)
```

Modelling Interaction Effect: Dummy Variables

```
Linear regression                               Number of obs     =     28,524
                                                F(4, 28519)       =    1987.84
                                                Prob > F          =     0.0000
                                                R-squared         =     0.2280
                                                Root MSE          =     .42014
```

ln_wage	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
black	-.0605195	.0063832	-9.48	0.000	-.0730309	-.0480081
not_smsa	-.1590936	.0064551	-24.65	0.000	-.1717459	-.1464413
smsa_black	-.0968763	.0117806	-8.22	0.000	-.1199668	-.0737858
grade	.0831061	.0011516	72.17	0.000	.0808489	.0853633
_cons	.7018273	.0152257	46.09	0.000	.6719842	.7316705

Statistical Inference from the Example

- We can see that grade or the level of education has a positive impact on wage. As education level increases by 1 unit, wage increases by around 8.31%, considering all are constant.

- It is clear that when the respondent is black, his(her) income is lower than the counterpart (i.e., white). On average, black employees earn 6.05% less than white employees, considering all are constant.

- Employees (black and white) not living in metropolitan areas earn 15.91% less than their counterpart, considering all are constant.

- Interestingly, Black employees who do not live in metropolitan areas earn 25.60% lower wage than Black employees living in metropolitan areas, considering all are constant.

  - $E(Wage|Black_i = 1, SMSA_i = 1) - E(Wage|Black_i = 1, SMSA_i = 0) = (\beta_1 + \beta_2 + \beta_3) - (\beta_1) = \beta_2 + \beta_3 = -25.60\%$

- This indicates locality affects wage differentials for black employees.

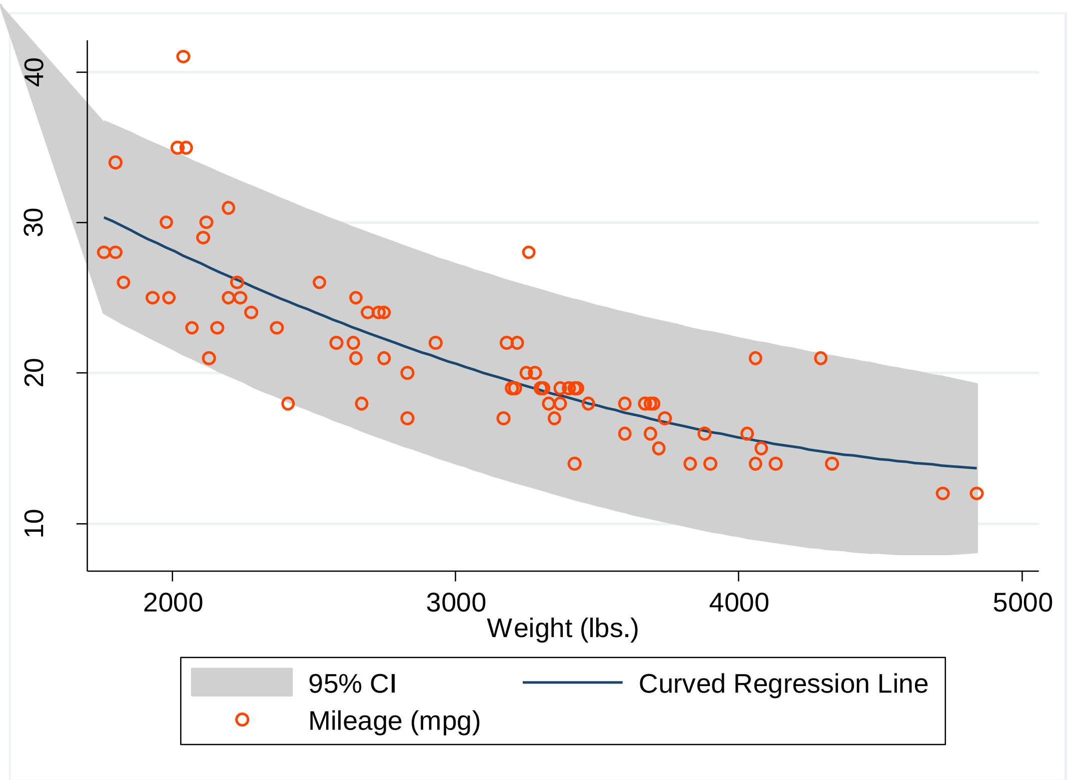
Non-Linearity in Causal Relationships

- Sometimes the relationship between the Y and X is not just a straight-line as we have seen so far.
- There may be some curvature in the relationship between the variables of interest. Think of a production function and the relationship between capital and output. Or, the EKC Hypothesis.
- Introduction of curvature makes the model specification a non-linear function
- By definition, a nonlinear function is a function which has a slope that is not constant.
- One way to approximate such a non-linear relationship mathematically is to model the relationship as a quadratic function.
- That is, we could model outcome variable Y as function of predictor X and the square of X.

A General Approach to Modeling Non-linearity

- Identify a possible non-linear relationship.
  - The best thing to do is to use economic theory and what we know about the application to suggest a possible non-linear relationship.
- Specify a non-linear function, and estimate its parameters by OLS/other estimators.
- Determine whether the nonlinear model improves upon a linear model.
  - We must determine empirically whether your non-linear model is appropriate.
- Plot the estimated non-linear regression function and check.

Visual Representation of Non-Linearity In Models



Non-linearity between mileage and weight of cars

Non-linearity between wage and total experience

STATA Commands:

sysuse auto
twoway qfitci mpg weight, stdr || scatter mpg weight

STATA Commands:

webuse regsmpl
twoway qfitci ln_wage ttl_exp , stdr || scatter ln_wage ttl_exp

Non-Linear Model Specifications

- General Expression for modelling non-linear relationship is:

$$Y = \alpha + \beta X + \gamma X^2 + \vartheta X^3 + \cdots + \nu X^n + u$$

- In different data setup, we can use our standard notations as subscripts for all variables:

$$Y_i = \alpha + \beta_1 X_{1\,i} + \beta_2 X_{1\,i}^2 + \beta_3 X_{1\,i}^3 + \cdots + \beta_n X_{n\,i}^n + u_i$$

$$Y_t = \alpha + \beta_1 X_{1\,t} + \beta_2 X_{1\,t}^2 + \beta_3 X_{1\,t}^3 + \cdots + \beta_n X_{n\,t}^n + u_t$$

$$Y_{it} = \alpha + \beta_1 X_{1\,it} + \beta_2 X_{1\,it}^2 + \beta_3 X_{1\,it}^3 + \cdots + \beta_n X_{n\,it}^n + u_{it}$$

- Apart from the estimation of the parameters, we can also find out where (i.e., the threshold) the curvature changes direction. This is widely known as turning point.
    - For second order and third order polynomials, the formula are: $\frac{-\beta_1}{2\beta_2}$ and $\frac{-\beta_2 \pm \sqrt{\beta_2^2 - 3\beta_1\beta_3}}{3\beta_3}$

Example of Non-Linear Model Estimation

- Let us take the following example:

$$lnWage_i = \alpha + \beta_1 lnAge_i + \beta_2 lnAge_i^2 + u_i$$

```
Linear regression                               Number of obs     =     28,510
                                                F(2, 28507)       =    1623.10
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0898
                                                Root MSE          =     .45614
```

ln_wage	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
ln_age	5.176267	.3304309	15.67	0.000	4.528607	5.823927
ln_age2	-.6879898	.0499695	-13.77	0.000	-.7859323	-.5900472
_cons	-7.902434	.5442279	-14.52	0.000	-8.969146	-6.835721

STATA Commands:

```
webuse regsmpl
g ln_age=log(age)
g ln_age2= (ln_age)^2
reg ln_wage ln_age ln_age2, vce(r)
```

Example of Non-Linear Model Estimation

We know turning point:

$$\frac{-\beta_1}{2\beta_2}$$

$$\frac{-5.176267}{2(-0.6879898\,)}$$

$$= 3.761878$$

$$= \exp(3.761878)$$

$$= 43.02915$$

So, when age reaches around 43 years (i.e., threshold), with higher age income will decline, ceteris paribus.