Notes 10: Introducing nonlinearities

ECO 231W - Undergraduate Econometrics

Prof. Carolina Caetano

1 Small math review

There is one important point about conditional expectations which you must know for this class. The rule can be quite complex mathematically, but it is very easy to understand in actual applications. Here is the principle, and don't be scared: if f is invertible,

$$\mathbb{E}[Y|f(X), g(X)] = \mathbb{E}[Y|X].$$

Wow, wait a minute, what? It means that if we condition on an invertible function of X, and then we may also add (or not) an arbitrary function of X, this is all the same as conditioning on X alone. Before you give up on this course, consider the following examples:

- $\mathbb{E}[grade|class^2] = \mathbb{E}[grade|class]$
- $\mathbb{E}[grade|class, class^2] = \mathbb{E}[grade|class]$
- $\mathbb{E}[grade|\log(class), class^2] = \mathbb{E}[grade|\log(class), class] = \mathbb{E}[grade|class]$

Think about those in English. In the first case, if we know $class^2$, isn't that the same as knowing $class$? Yes, it is. In the second case, if we know $class$, does $class^2$ add any information? No. In the third case, if we know $class^2$, then we know $class$ (which explains the first equality), but if we know $class$, then what new information is $\log(class)$ bringing? None.

2 What do we mean by nonlinearities?

First, I want you to realize that we are not going to abandon the linear model. The linear model:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u,$$

with $\mathbb{E}[u|x_1, \ldots, x_k] = 0$ is called linear because it is "linear in parameters." What does this mean? It means that the expected outcome is equal to a constant (β_0) plus a constant times a variable ($\beta_1 x_1$) plus ... plus a final constant times a final variable ($\beta_k x_k$) plus u. We never said what those variables should be. The variable *income* is just as much of a variable as $\log(income)$, as e^{income}, and so on. Nobody said that you were obligated to use the variables just as they came in the data set. It is fine to modify them if it suits the relationships. This doesn't affect the linearity of the model.

So, when I say that we will incorporate nonlinearities into the model, I don't mean to say that we are leaving the Linear Model at all. For example, consider the model

$$wage = \beta_0 + \beta_1 education + \beta_2 experience + u,$$

with $\mathbb{E}[u|education, experience] = 0$. This model is quite bad. Why? Because the relation between wages and experience is not linear at all. When we move to the highest levels of experience, generally we are talking about low education professions, which are not very well remunerated, were people can start working very early in life, and where their salaries don't necessarily go up much: housekeepers, construction workers, etc. This linear model implies that as experience increases, wages will be expected to increase exactly the same amount, no matter what was the original experience. What I am saying is that this is not true. As experience increases, the wages will likely go initially up, and eventually go down. So, a much better model is (in the sense that it is more likely to be true)

There are a few things that you should notice:

- ____
  ____
  ____

- ____

  ____

  ____

  ____

- ____

  ____

3 Interpreting coefficients

Using transformations of variables can affect the interpretation of the coefficients. All the operations we were doing before in order to understand the meaning of the βs were specific for linear relations between variables. Let's continue looking at the model

$$wage = \beta_0 + \beta_1 education + \beta_2 experience + \beta_3 experience^2 + u$$

with $\mathbb{E}[u|education, experience] = 0$.

When the relation between the outcome ($wage$) and the variable is linear, as is the case with education, the interpretation of the corresponding β remains the same.

- What is β_1? We first look at the equations:

  and

so

In English: ____

____

____

The problem is the interpretation of the coefficients of the variables that have non-linear relationships with the outcome, as is the case with *experience*. Often, there is no economic interpretation at all, in the sense that we cannot come up with a neat understanding of the intuitive meaning of that β. I think that this is the case with β_2 and β_3 in this model. However, for certain forms on lon-linearities, the interpretation can be very interesting. Let's take a look at the next type of non-linearity.

4 Models with logs

It is quite common to see models where variables are substituted by logs. Depending on which variable is a log, the interpretation can change. Consider the model:

$$grade = \beta_0 + \beta_1 \log(class) + \beta_2 OH + u$$

where $\mathbb{E}[u|class, OH] = 0$.

- What is β_1? The trick is to understand the properties of the log function. What I will do next can be easily understood if you remember derivatives, especially Taylor expansions. However, you don't need to. Here is the property you need to know:

$$log(x') - \log(x) \approx \frac{x' - x}{x}$$

What is the meaning of $\frac{x'-x}{x}$? ____

____

In other words, when we use a log in a model, the coefficients are now interpreted in terms of proportional variation, instead of the actual variation. To see this, observe that when $class = c$ **(space on the next page)**, and now let's vary *class*. The new level is $class = c'$. Leaving everything else constant,

and subtracting the first from the second:

So

In English: ____

____

____

- What is β_0? Let's look at the equation.

Now, when you go home, experiment interpreting the different coefficients when the outcome variable is the one with logs. Try the following exercises:

Exercise 4.1. *Interpret the coefficients in the model*

$$\log(grade) = \beta_0 + \beta_1 class + \beta_2 OH + u$$

where $\mathbb{E}[u|class, OH] = 0$.

Exercise 4.2. *Interpret the coefficients in the model*

$$\log(grade) = \beta_0 + \beta_1 \log(class) + \beta_2 OH + u$$

where $\mathbb{E}[u|class, OH] = 0$.

5 Models with interactions

A very common form on non-linearity is adding products of the variables. For example,

The point of this style of model is to pick up the interaction between the variables. The idea is that *class* and *OH* may affect *grade*, yes, but they can also interact. The idea is that one can have a reward of going to classes, a reward of going to office hours, and a special reward of the interaction, an extra something that only comes if you do both. In this example, this interaction has little intuition, but this way of thinking is quite useful when using qualitative data. For example, suppose that our data set has information on whether the student is smart or not. We can build the dummy *smart*, and write the model

$$grade = \beta_0 + \beta_1 class + \beta_2 OH + \beta_3 smart + u$$

where $\mathbb{E}[u|class, OH, smart] = 0$. This model allows that the classes and office ours influence the grade, as well as whether the student is smart or not. However, we can think that smart students will take different advantage from class compared to those who are not smart. The point is that this model gave the same reward for classes, independent of the smartness level of the student. I think that this is unlikely to be true. A better model is

$$grade = \beta_0 + \beta_1 class + \beta_2 OH + \beta_3 smart + \beta_4 smart \cdot class + u$$

$\mathbb{E}[u|class, OH, smart] = 0$.

In this model, the interpretation of β_0 and β_2 is the same you would expect (check this). However, all the other coefficients change. If you are faced with a situation where there are interactions, I suggest that you try to open all the categories until you get the pattern. In this case, the qualitative variable is *smart*. Hence,

- When $smart = 0$,

- When $smart = 1$,

You have to see this as if there were two models, one for smart people, and the other for those that aren't. The model for those that aren't has no complications. The other one we can figure out later. So, if you do the math, you will see that

$$\beta_1 = \mathbb{E}[grade|class = c + 1, OH, smart = 0] - \mathbb{E}[grade|class = c, OH, smart = 0].$$

In English, β_1 is the expected change in grade caused by one extra class among not smart people, leaving everything else constant.

Now, what is β_4? If you do the math, you will see that

$$\beta_4 = \mathbb{E}[grade|class = c + 1, OH, smart = 1] - \mathbb{E}[grade|class = c, OH, smart = 1] - \beta_1.$$

In English, β_4 is the expected extra premium in the grade for being smart caused by one extra class, leaving everything else constant.

Finally, what is β_3? If you do the math, you will see that

$$\beta_3 = \mathbb{E}[grade|class = 0, OH, smart = 1] - \mathbb{E}[grade|class = 0, OH, smart = 0].$$

In English, β_3 is the expected difference in grade between smart and not smart students which do not go to class, leaving everything else constant.

5.1 Additional examples

This section is not part of the class, but it provides more examples of how to interpret interactions, which may help you understand the pattern. I suggest that you try those on your own, and then use these examples to check your answer.

Suppose that we have a second qualitative variable, which denotes whether the student has a large network of friends taking the course. The model could be

$$grade = \beta_0 + \beta_1 class + \beta_2 OH + \beta_3 smart + \beta_4 connected + u$$

where $\mathbb{E}[u|class, OH, smart, connected] = 0$. This model says that being smart and connected can matter. In fact, one would get the rewards of being smart (β_3) no matter whether the student is well connected or not, for example because a smart student gets more out of classes, and out of studying, than other students. One can get rewards of being connected independent of being smart or not (β_4). For example, independent of whether he or she is smart or not, the student can enjoy coming to class if their friends are also in it. The student can have more incentive to study, the student can get access to homeworks and practice exams from others. However, we could think that being smart has special rewards when one is also connected, so that the combination of the two things is special on its own. For example, a connected student receives homeworks and practice exams, but if he or she is also smart, they can understand and improve on the answers received, so that the final reward is not simply the sum of the reward of being smart, and that of being connected, but even higher. The model should then be

$$grade = \beta_0 + \beta_1 class + \beta_2 OH + \beta_3 smart + \beta_4 connected + \beta_5 smart \cdot connected + u,$$

where $\mathbb{E}[u|class, OH, smart, connected] = 0$.

The coefficients β_0, β_1 and β_2 have the same interpretation you would expect (and you should check that you would know how to show this). However, the interpretation of β_3, β_4 and β_5 is more complicated. In the same way we did in class, we begin by understanding what the model really is saying for each possible category. We have two binary qualities, which make four possibilities:

- What is the expected grade of a smart connected person?

$$\mathbb{E}[grade|class, OH, smart = 1, connected = 1] = \beta_0 + \beta_1 class + \beta_2 OH + \beta_3 + \beta_4 + \beta_5$$

- What is the expected grade of a smart unconnected person?

$$\mathbb{E}[grade|class, OH, smart = 1, connected = 0] = \beta_0 + \beta_1 class + \beta_2 OH + \beta_3$$

- What is the expected grade of a not smart, connected person?

$$\mathbb{E}[grade|class, OH, smart = 0, connected = 1] = \beta_0 + \beta_1 class + \beta_2 OH + \beta_4$$

- What is the expected grade of a not smart, unconnected person?

$$\mathbb{E}[grade|class, OH, smart = 0, connected = 0] = \beta_0 + \beta_1 class + \beta_2 OH$$

If you do the operations, you will see that

$$\begin{aligned}\beta_3 = \mathbb{E}&[grade|class, OH, smart = 1, connected = 0] \\ &- \mathbb{E}[grade|class, OH, smart = 0, connected = 0]\end{aligned}$$

(in English, β_3 is the reward of being smart versus not smart among unconnected people, leaving everything else constant.)

$$\begin{aligned}\beta_4 = \mathbb{E}&[grade|class, OH, smart = 0, connected = 1] \\ &- \mathbb{E}[grade|class, OH, smart = 0, connected = 0]\end{aligned}$$

(in English, β_4 is the reward of being connected versus unconnected among not smart people, leaving everything else constant.)

$$\begin{aligned}\beta_5 = &\Big[\mathbb{E}[grade|class, OH, smart = 1, connected = 1] \\ &\quad - \mathbb{E}[grade|class, OH, smart = 0, connected = 0]\Big] \\ &- \beta_3 - \beta_4\end{aligned}$$

(in English, β_5 is the reward of being smart and connected versus being neither, net of the rewards of only being either smart or connected, leaving everything else constant.)

Another example

Now let's try the next degree of complexity:

$$grade = \beta_0 + \beta_1 \log(class) + \beta_2 OH + \beta_3 smart + \beta_4 smart \cdot \log(class) + u,$$

where $\mathbb{E}[u|class, OH, smart] = 0$.

The coefficients β_2 has the same interpretation you would expect (and you should check that you would know how to show this). However, the interpretation of β_0, β_1, β_3, and β_4 change. We begin with the easiest:

- What is β_0? If you do the math, you will see that

$$\beta_0 = \mathbb{E}[grade|class = 1, OH = 0, smart = 0]$$

  (in English, β_0 is the expected grade of a student that goes to one class, attends no office hour, and is not smart.

The interpretation of β_0 above was affected only because of the presence of the log. There is nothing about the interaction that made any difference. The interpretation of the coefficients of the terms involved in the interactions is mode complex. We start again in the same way as before, looking at the model implications for each category.

pattern. In this case, the qualitative variable is *smart*. Hence,

- When $smart = 0$,

$$\mathbb{E}[grade|class, OH, smart = 0] = \beta_0 + \beta_1 \log(class) + \beta_2 OH.$$

- When $smart = 1$,

$$\begin{aligned}\mathbb{E}[grade|class, OH, smart = 1] &= \beta_0 + \beta_1 \log(class) + \beta_2 OH + \beta_3 + \beta_4 \log(class) \\ &= \beta_0 + (\beta_1 + \beta_4) \log(class) + \beta_2 OH + \beta_3\end{aligned}$$

You have to see this as if there were two models, one for smart people, and the other for those that aren't. The model for those that aren't has no complications. The other one we can figure out later. The only novelty here is that we have logs, but we know what logs mean, right? They mean proportional changes. So, if you do the math, you will see that

$$\beta_1 = \frac{\mathbb{E}[grade|class = c', OH, smart = 0] - \mathbb{E}[grade|class = c, OH, smart = 0]}{\frac{c'-c}{c}}.$$

In English, β_1 is the expected corresponding change in grade caused by a proportional change in class among not smart people, leaving everything else constant.

Now, what is β_4? If you do the math, you will see that

$$\beta_4 = \frac{\mathbb{E}[grade|class = c+1, OH, smart = 1] - \mathbb{E}[grade|class = c, OH, smart = 1]}{\frac{c'-c}{c}} - \beta_1.$$

In English, β_4 is the expected corresponding extra premium in the grade for being smart caused by a proportional change in class, leaving everything else constant.

Finally, what is β_3? If you do the math, you will see that

$$\beta_3 = \mathbb{E}[grade|class = 1, OH, smart = 1] - \mathbb{E}[grade|class = 1, OH, smart = 0].$$

In English, β_3 is the expected difference in grade between smart and not smart students which go to one class, leaving everything else constant.