

Topic 2 Introductory Note: Measuring non-linear relationships

These notes are provided as a supplement to the lecture slides in BSM946 and are not expected to be used as your sole understanding of this topic. The main arguments are presented in the lecture, with extra notes here and the key reading to ensure your understanding of the topic is complete.

Throughout this course we will be interested in finding relationships between variables. These will not always be nice linear relationships as in the real world this does not happen. Thus we need to develop some useful tools to effectively model relationships that are non-linear. These notes will discuss two simple methods, logs and squared terms.

Part 1: Logs

Why do we use logs?

Not all variables will have linear relationships and often they take more curved, exponential looking forms. By taking logs we are able to transform a given variable to a linear form in a lot of cases, allowing us to fit a linear regression better. There are some useful notes on this in the following video:

<https://www.youtube.com/watch?v=HlcqQhn3vSM>

Log-Linear Models

If you use natural log values for your dependent variable (Y) and keep your independent variables (X) in their original scale, the econometric specification is called a *log-linear model*. These models are typically used when you think the variables may have an exponential growth relationship.

$$\text{Log}(Y) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

The interpretation of the coefficients changes slightly by doing this however!

Rather than a 1 unit increase in X_1 causing a β_1 increase in Y , a one unit increase in X_1 now causes a β_1 percent increase in Y . This is quite an important distinction as a 1 unit increase in a variable such as GDP is small, but a 1% increase in GDP is very large!

Linear-Log Models

We can also use the reverse of the last equation and log the independent variables but not the dependent variable:

$$Y = \alpha + \beta_1 \log(X_1) + \beta_2 \log(X_2) + \varepsilon$$

Rather than a 1 unit increase in X_1 causing a β_1 increase in Y , a one per cent increase in X now causes a β_1 unit increase in Y .

Log-Log Models

Finally, we can have what we call a log-log model (or a double log model), with both the dependent and independent variables being in a logged form:

$$\text{Log}(Y) = \alpha + \beta_1 \log(X_1) + \beta_2 \log(X_2) + \varepsilon$$

Rather than a 1 unit increase in X_1 causing a β_1 increase in Y , a one per cent increase in X_1 now causes a β_1 per cent increase in Y . These are used quite a lot in economics as it is actually a straightforward elasticity and thus its interpretation is something we are fairly used to.

Final Point

Just be careful what form your variables are in before you start interpreting coefficients! Logging variables does often help us fit models more accurately however so do think about transforming your variables where appropriate.

Part 2: Squared terms

Why do we see squares in regressions?

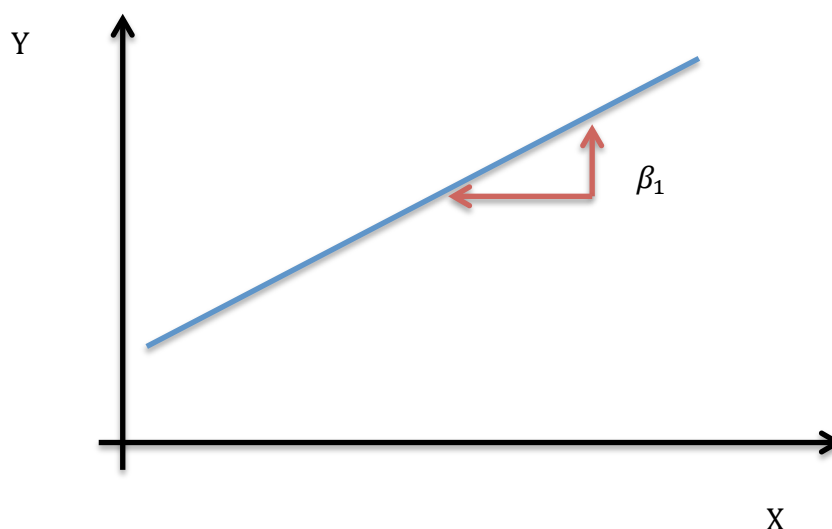
Often when trying to model relationships in econometrics you will see squared terms in our regression models. Why is this? Well the basic idea is that they again allow us to take into account some of the non-linearity that may exist in our relationship so we can better measure it. Let's work through a quick example.

Consider we are trying to model a relationship between Y and X such:

$$Y = \alpha + \beta_1 X_1 + \varepsilon$$

Where we can determine how much X influences Y by taking the first partial derivative of Y with respect to X , which should give us β_1 .

$$\frac{dY}{dX_1} = \beta_1$$



Here we are saying that the slope of the line, and thus the relationship between Y and X does not depend upon the value of X. The slope is the same at all points on the line, and is determined by beta.

This is not always the case and that is why we sometimes include a squared version of our independent variable. Consider now the equation:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_1^2 + \epsilon$$

If we take the first derivative now we get:

$$\frac{dY}{dX_1} = \beta_1 + 2 * \beta_2 X_1$$

So the relationship between Y and X is dependent upon both β_1 as well as the $\beta_2 * X_1$. i.e. the relationship is not always the same, it depends upon the value of X as well as beta.

We can see this below, that as we move along the X axis, the relationship between X and Y is not constant, and thus our line is non-linear. Including our square term allows us to control for this. It allows us to determine if there is a positive or negative relationship between X and Y (which will be shown by the β_1 coefficient), and whether this relationship is increasing or decreasing as we increase X (which will be shown by the β_2 coefficient). The below equation would give us a value of $\beta_1 > 0$, as there is a positive relationship between X and Y (as we increase X, Y gets larger) but with $\beta_2 < 0$, showing that as we increase X the change in Y gets smaller and smaller!

Final Point

It is worth playing around with your specification a little to ensure you are capturing the true relationship as best you can. It often helps to graph variables against each other to aid this process!

