L4: Extensions of the two-variable linear regression model

Feng Li
m@feng.li

School of Statistics and Mathematics
Central University of Finance and Economics

Revision: October 7, 2014

What we have learned last time...

1. Normal assumptions of u_i
2. Constructing intervals, Hypothesis testing
3. Normality tests

Today we are going to learn...

1. Regression through the origin

2. Scaling and units of measurement

3. Regression on standardized variables

4. Other models

Regression through the origin

↪ The estimation

1. Assume the population regression function of the form

$$Y_i = \beta_2 X_i + u_i$$

2. The sample regression function is

$$Y_i = \beta_2 X_i + \hat{u}_i$$

3. Following the same procedure of OLS we have

$$\hat{\beta}_2 = \frac{\sum X_i Y_i}{\sum X_i^2}, \qquad var(\hat{\beta}_2) = \frac{\sigma^2}{\sum X_i^2}, \qquad \hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-1}$$

4. Comparing them with the model with intercept. **What can you find**?

Regression through the origin
↪ Use with caution

1. The r^2 defined for linear with intercept is not applied here. We defined that

$$r^2 = 1 - RSS/TSS$$

where

$$RSS = \sum u_i^2 = \sum Y_i^2 - \hat{\beta}_2^2 \sum X_i^2, \qquad TSS = \sum (Y_i - \bar{Y})^2 = \sum Y_i^2 - n\bar{Y}^2$$

in the no intercept model. There is **no guarantee** that $RSS \leqslant TSS$. Hence $r^2 < 0$ is possible.

2. Instead, people use **raw** r^2 in this model,

$$raw\ r^2 = \frac{(\sum X_i Y_i)^2}{\sum X_i^2 \sum Y_i^2}$$

3. The **raw** r^2 is not comparable with r^2.
4. Specification error may occur if you insist use this model. **See Chapter 7**.

Scaling and units of measurement
↪ The model

1. Suppose that you want to model

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i$$

2. While you observe the data $Y_i^* = w_1 Y_i$ and $X_i^* = w_2 X_i$
3. So you have to consider instead

$$Y_i^* = \hat{\beta}_1^* + \hat{\beta}_2^* X_i^* + \hat{u}_i^*$$

   where $\hat{u}_i^* = w_1 \hat{u}_i$ (**why?**)
4. It is important to know that the two model are essentially the same because scaling the units will not change the data and the properties of the OLS not changed.

Scaling and units of measurement
↪ The estimation

1. It is easy to obtain the following occlusions (**homework!**)
   1. $\hat{\beta}_2^* = (\frac{w_1}{w_2})\hat{\beta}_2$
   2. $\hat{\beta}_1^* = w_1\hat{\beta}_1$
   3. $(\hat{\sigma}^*)^2 = w_1^2\hat{\sigma}^2$
   4. $var(\hat{\beta}_1^*) = w_1^2 var(\hat{\beta}_1)$
   5. $var(\hat{\beta}_2^*) = (\frac{w_1}{w_2})^2 var(\hat{\beta}_2)$
   6. $r_{xy}^2 = r_{x*y*}^2$
2. Example:

$$\hat{\beta}_2^* = \frac{\sum x_i^* y_i^*}{\sum (x_i^*)^2} = \frac{\sum (X_i^* - \bar{X}^*)(Y_i^* - \bar{Y}^*)}{\sum (X_i^* - \bar{X}^*)^2} = \frac{\sum (w_2 X_i - w_2 \bar{X})(w_1 Y_i - w_1 \bar{Y})}{\sum (w_2 X_i - w_2 \bar{X})^2}$$

$$= \frac{w_2 w_1}{w_2^2} \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{w_1}{w_2}\hat{\beta}_2$$

Regression on standardized variables
↪ The model

1. The same model as before, but the transformation are not

$$Y_i^* = \frac{Y_i - \bar{Y}}{S_Y}, \qquad X_i^* = \frac{X_i - \bar{X}}{S_X}$$

2. The model is now as

$$Y_i^* = \beta_1^* + \beta_2^* X_i^* + u_i^* = \beta_2^* X_i^* + u_i^*$$

which goes through the origin.

3. The new coefficient

$$\hat{\beta}_2^* = \hat{\beta}_2 \frac{S_X}{S_Y}$$

The log-linear model

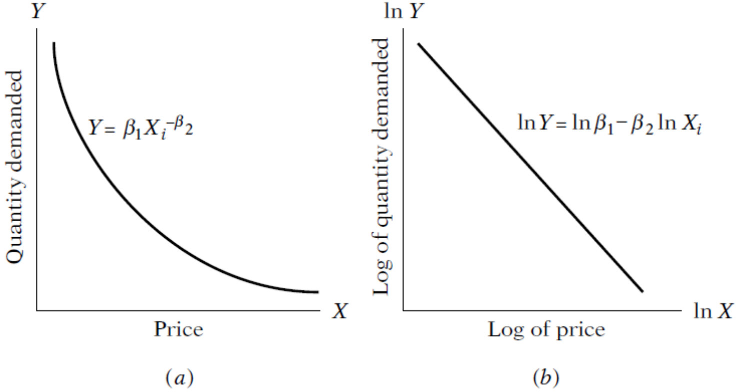
1. Consider the **exponential regression model**

$$Y_i = \beta_1 X_i^{\beta_2} e^{u_i}$$

2. We express it as

$$\ln Y_i = \ln\beta_1 + \beta_2 \ln X_i + u_i$$

so that the linear model estimation can be used.



The log-lin model

1. Consider the model
$$Y_t = Y_0(1+r)^t$$
2. We express it as
$$\ln Y_t = \ln Y_0 + [\ln(1+r)]t$$
3. Let $\beta_1 = \ln Y_0$ and $\beta_2 = \ln(1+r)$ we have the following model
$$\ln Y_t = \beta_1 + \beta_2 t + u_t$$

The lin-log model

1. The lin-log model
$$Y_i = \beta_1 + \beta_2 \ln X_i + u_t$$
where $\hat{\beta}_2 = \frac{\Delta Y}{\Delta X / X}$ and Δ means very small change.
2. Thus $\Delta Y = \hat{\beta}_2 \cdot (\Delta X / X)$
3. The interpretation of this model:
If $\Delta X / X$ changes by 1 percent, the absolute change in Y is $0.01\beta_2$

Summary of models

Model	Equation	Slope $\left(=\frac{dY}{dX}\right)$	Elasticity $\left(=\frac{dY}{dX}\frac{X}{Y}\right)$
Linear	$Y = \beta_1 + \beta_2 X$	β_2	$\beta_2\left(\frac{X}{Y}\right)^*$
Log–linear	$\ln Y = \beta_1 + \beta_2 \ln X$	$\beta_2\left(\frac{Y}{X}\right)$	β_2
Log–lin	$\ln Y = \beta_1 + \beta_2 X$	$\beta_2 (Y)$	$\beta_2 (X)^*$
Lin–log	$Y = \beta_1 + \beta_2 \ln X$	$\beta_2\left(\frac{1}{X}\right)$	$\beta_2\left(\frac{1}{Y}\right)^*$
Reciprocal	$Y = \beta_1 + \beta_2\left(\frac{1}{X}\right)$	$-\beta_2\left(\frac{1}{X^2}\right)$	$-\beta_2\left(\frac{1}{XY}\right)^*$
Log reciprocal	$\ln Y = \beta_1 - \beta_2\left(\frac{1}{X}\right)$	$\beta_2\left(\frac{Y}{X^2}\right)$	$\beta_2\left(\frac{1}{X}\right)^*$

Note: * indicates that the elasticity is variable, depending on the value taken by X or Y or both. When no X and Y values are specified, in practice, very often these elasticities are measured at the mean values of these variables, namely, $\bar{X}$ and $\bar{Y}$.

Take home questions

1. Verify the results in **slide 7**.
2. Think about when you make log transformation, will it change the distribution of the error term?
3. Exercises (S2): **5.1, 5.5, 5.8, 5.14, 5.15, 6.1, 6.2, 6.7, 6.13, 6.14**