Econometrics

Chapter 7: Multiple Regression: Further Issues

In Choi

Sogang University

Redefining Variables

- Changing the scale of the y variable will lead to a corresponding change in the scale of the coefficients and standard errors, so no change in the significance or interpretation

Illustration Consider the original model

$$y_t = \beta_0 + \beta_1 x_t + u_t.$$

Let the OLS estimator of β_1 be $\hat{\beta}_1$. Suppose that cy_t is used instead of y_t and let the resulting OLS estimator be $\bar{\beta}_1$. Then, $\bar{\beta}_1 = c\hat{\beta}_1$ and $Var(\bar{\beta}_1) = c^2 Var(\hat{\beta}_1)$. Thus, the t-ratio for the null hypothesis $H_0 : \beta_1 = 0$ does not change when cy_t is used.

Redefining Variables

- Changing the scale of one x variable will lead to a change in the scale of that coefficient and standard error, so no change in the significance or interpretation

Illustration Consider the original model

$$y_t = \beta_0 + \beta_1 x_t + u_t.$$

Let the OLS estimator of β_1 be $\hat{\beta}_1$. Suppose that cx_t is used instead of x_t and let the resulting OLS estimator be $\bar{\beta}_1$. Then, $\bar{\beta}_1 = \frac{1}{c}\hat{\beta}_1$ and $Var(\bar{\beta}_1) = \frac{1}{c^2}Var(\hat{\beta}_1)$. Thus, the t-ratio for the null hypothesis $H_0 : \beta_1 = 0$ does not change when cx_t is used.

Beta Coefficients

- Occasional we see reference to a "standardized coefficient" or "beta coefficient" which has a specific meaning.
- Idea is to replace y and each x variable with a standardized version – i.e. subtract mean and divide by standard deviation
- Coefficient reflects standard deviation of y for a one standard deviation change in x.

Functional Form

- OLS can be used for relationships that are not strictly linear in and y by using nonlinear functions of x and y.
  - Can take the natural log of x, y or both
  - Can use quadratic forms of x
  - Can use interactions of x variables

- If the model is

$$ln(y_t) = \beta_0 + \beta_1 ln(x_t) + u_t,$$

β_1 is the elasticity of y with respect to x.

- If the model is

$$ln(y_t) = \beta_0 + \beta_1 x_t + u_t,$$

β_1 is approximately the growth rate of y given a 1 unit change in x. Note that

$$\beta_1 = \frac{\partial \ln(y)}{\partial x} = \frac{\frac{\partial y}{y}}{\partial x}.$$

Example

If $x_t = t$ and y_t is the annual GDP, β_1 is the average, annual GDP growth rate over the sampling period.

- If the model is

$$y_t = \beta_0 + \beta_1 ln(x_t) + u_t,$$

$$\beta_1/100 = \frac{\partial y}{100 \times \partial \ln(x)} = \frac{\partial y}{100 \times \frac{\partial x}{x}}.$$

That is, $\beta_1/100$ is approximately the change in y for an 1% change in x. In other word, β_1 is approximately the change in y for an 100% change in x.

- Log models are invariant to the scale of the variables since they measure percent changes.

Illustration Consider the original model

$$\ln(y_t) = \beta_0 + \beta_1 \ln(x_t) + u_t.$$

Suppose that $y_t \to y_t^* = c_y y_t$ and $x_t \to x_t^* = c_x x_t$. Then the model becomes

$$\ln(y_t^*/c_y) = \beta_0 + \beta_1 \ln(x_t^*/c_x) + u_t$$

or

$$\begin{aligned} \ln(y_t^*) &= (\ln(c_y) - \beta_1 \ln c_x + \beta_0) + \beta_1 \ln(x_t^*) + u_t \\ &= \beta_0^* + \beta_1 \ln(x_t^*) + u_t. \end{aligned}$$

Thus, the coefficient β_1 is interpreted in the same way as in the original model.

Functional Form

Why use log models?

- They give a direct estimate of elasticity.
- For models with $y > 0$, the conditional distribution is often heteroskedastic or skewed, while $ln(y)$ is much less so.
- The distribution of $ln(y)$ is more narrow, limiting the effect of outliers.

Functional Form
Some Rules of Thumb

- What types of variables are often used in log form?
  - Dollar amounts that must be positive
  - Very large variables, such as population
- What types of variables are often used in level form?
  - Variables measured in years
  - Variables that are a proportion or percent

- For a model of the form

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 x_t^2 + u_t,$$

we can't interpret β_1 alone as measuring the change in y with respect to x, we need to take into account β_2 as well. Since $\partial y = (\beta_1 + 2\beta_2 x)\partial x$,

$$\frac{\partial y}{\partial x} = \beta_1 + 2\beta_2 x.$$

Functional Form

More on Quadratic Models

- Suppose that the coefficient on x is positive and the coefficient on x^2 is negative.
- Then y is increasing in x at first, but will eventually turn around and be decreasing in x.
- For $\hat{\beta}_1 > 0$ and $\hat{\beta}_2 < 0$, the turning point will be

$$x^* = \left|\hat{\beta}_1 / 2\hat{\beta}_2\right|$$

Example

y_t : log wage, x_t : number of years in school

Example

y_t : Gini coefficient, x_t :log per capita GDP. As an economy grows, there may be higher level of income inequality. However, this may dissipate gradually as the economy grows further and becomes mature.

- Suppose that the coefficient on x is negative and the coefficient on x^2 is positive.
- Then y is decreasing in x at first, but will eventually turn around and be increasing in x.
- For $\hat{\beta}_1 < 0$ and $\hat{\beta}_2 > 0$, the turning point will be

$$x^* = \left|\hat{\beta}_1 / 2\hat{\beta}_2\right|.$$

- For a model of the form

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{1t} x_{2t} + u_t,$$

we can't interpret β_1 alone as measuring the change in y with respect to x_1, we need to take into account β_3 as well, since

$$\frac{\partial y}{\partial x_1} = \beta_1 + \beta_3 x_2.$$

Example

$$D = \beta_1 + \beta_2 S + \beta_3 W + \beta_4 SW + \varepsilon$$

D : breaking distance

W : road wetness

S : speed

$$\frac{\partial D}{\partial S} = \beta_2 + \beta_4 W$$

If $\beta_4 > 0$, the marginal effect of higher speed on breaking distance is increased when the road is wetter.

Predictions

- Suppose we want to use our estimates to obtain a specific prediction.

Example

What is the expected wage of a married, male worker with 12 years in school and 10 years' work experience in banking industry?

- An estimate of $E(y^o|x_1 = c_1, \ldots, x_k = c_k)$ is

$$q^o = \beta_0 + \beta_1 c_1 + \ldots + \beta_k c_k. \quad (1)$$

  This is an estimate of y when x's take some particular values.

- Using this, we obtain

$$y_i = q^o + \beta_1(x_{1i} - c_1) + \ldots + \beta_k(x_{ki} - c_k) + u_i. \quad (2)$$

  So, if you regress y_i on $(x_{1i} - c_1), \ldots, (x_{ki} - c_k)$, the intercept will give the predicted value and its standard error.

Predictions

- Let the OLS estimators from equation (2) be $\hat{q}^o, \hat{\beta}_1, ..., \hat{\beta}_k$. The estimators $\hat{\beta}_1, ..., \hat{\beta}_k$ are the same as the corresponding OLS estimators obtained from the regression equation

$$y_i = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_k x_{ki} + u_i. \tag{3}$$

- We have

$$\begin{aligned} \hat{q}^o &= \bar{y} - \hat{\beta}_1(\bar{x}_1 - c_1) - ... - \hat{\beta}_k(\bar{x}_k - c_k) \\ &= \bar{y} - \hat{\beta}_1\bar{x}_1 - ... - \hat{\beta}_k\bar{x}_k + \hat{\beta}_1 c_1 + ... + \hat{\beta}_k c_k \\ &= \hat{\beta}_0 + \hat{\beta}_1 c_1 + ... + \hat{\beta}_k c_k, \end{aligned}$$

  where $\hat{\beta}_0$ is the OLS estimator of β_0 from equation (3). Thus, $\hat{q}^o$ is the same as its estimator using the OLS estimators from the original linear model (3) along with equation (1).

- Note that the standard error will be smallest when the c's equal the means of the x's.

Predictions

Illustration In the case of simple linear regression,

$$Var(\hat{q}^o) = \frac{\sigma^2}{n} \frac{\sum_{i=1}^n (x_i - c)^2}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

This is minimized when $c = \bar{x}$.

- Let the prediction error be

$$\begin{aligned} \hat{e}^o &= y^o - \hat{q}^o \\ &= \beta_0 + \beta_1 c_1 + \ldots + \beta_k c_k + u^o - \hat{q}^o \\ &= u^o - (\hat{\beta}_0 - \beta_0) - \sum_{i=1}^k (\hat{\beta}_i - \beta_i) c_i. \end{aligned}$$

Then

$$E(\hat{e}^o) = 0$$

and

$$Var(\hat{e}^0) = Var(\hat{q}^o) + Var(u^o) = Var(\hat{q}^o) + \sigma^2.$$

Predictions

- $Var(\hat{q}^o)$ is estimated by regression (2). Thus, replacing σ^2 with its estimator s^2, we obtain the standard error of $\hat{e}^0$ ($se(\hat{e}^0)$).
- Since

$$\frac{\hat{e}^0}{se(\hat{e}^0)} \sim t_{n-k-1}$$

under a normality assumption, a 95% prediction interval for y^0 is

$$\hat{q}^o \pm t_{0.025} \times se(\hat{e}^0).$$

- Usually, the estimate of s^2 is much larger than the variance of the prediction $Var(\hat{q}^o)$.