Empirical exercise – Inference – Delta method

1. Aim of the exercise

The aim of this exercise is to learn to carry out a hypothesis test on a nonlinear function of the population coefficient.

The theoretical context, particular to this empirical exercise, is as follows. Consider two numbers 41 and 40. The 'logarithmic change' is $\ln(41) - \ln(40) = 0.02469$ rounded to the fifth decimal. The 'proportionate change', or the 'relative change' is $(41 - 40)/40 = 0.025$. The two quantities are very close to each other. This shows that for small changes the logarithmic change closely approximates the proportionate change. If we multiply the proportionate change by 100 we obtain the 'percentage change' which is 2.5%. Now consider the numbers 60 and 40. The logarithmic change is $\ln(60) - \ln(40) = 0.40546$ rounded to the fifth decimal, whereas the proportionate change is $(60-40)/40 = 0.500$. This shows that for large changes the logarithmic change is not a good approximation of the proportionate change.

Consider the log-linear regression model that takes the form $\ln(y) = \beta_1 x_1 + u$. For a small change in x_1, the 'logarithmic change' in y will closely approximate the exact proportionate change in y. In this case we can check how the marginal change $\partial ln(y)$ is affected by the marginal change $\partial ln(x_1)$. This is given by β_1. For a large change in x_1, the approximation will be worse, as discussed above. In this case β_1 is not satisfactory. We should like to check how the exact proportionate change in y is determined by a discrete change in x_1. Start by considering $\Delta \ln(y) = \ln(y_1) - \ln(y_0) = \beta_1 \Delta x_1$. Taking the exponent of both sides, we obtain $\exp(\ln(y_1) - \ln(y_0)) = \exp(\ln(y_1/y_0)) = y_1/y_0 = \exp(\beta_1 \Delta x_1)$. Subtracting 1 from each side of $y_1/y_0 = \exp(\beta_1 \Delta x_1)$ gives the exact proportionate change in y as $(y_1 - y_0)/y_0 = \exp(\beta_1 \Delta x_1) - 1$. This means that, in the log-linear regression model, if we are interested in knowing the exact proportionate change in y for some discrete change in x_1, we should consider the expression $\exp(\beta_1 \Delta x_1) - 1$ and not β_1. Given the discrete effect, we may then want to test whether this discrete population effect is equal to some hypothesised value. We can consider a hypothesis test to answer this question. However, note that this discrete change is non-linear in the coefficient of x_1. We need to calculate the standard error of this expression to conduct a test on the discrete effect. How do we obtain this standard error?

The empirical context is as follows. Suppose that we analyse the factors that determine the test scores of students in high school. We have data on test scores, and on the average income of the district where the student is from, among a set of other characteristics. We are particularly interested in the effect of average income in a district on the test scores. To investigate this, we estimate the log-linear model that takes the form $\ln(testscr) = \beta_1 avginc + u$. Suppose we are interested in the effect of increasing the average income by 5 units (that is by \$5000). We can estimate the exact proportionate change in test scores as $\exp(\hat{\beta}_1 * 5) - 1$. We would ask whether this discrete effect is significantly different from, say, zero. We can consider a t test to answer this question but for this we need the standard error of the stated expression. How do we obtain this standard error?

2. Load the data

Load the data on test scores in mat format into MATLAB.

```
clear;
load 'M:\exercise.mat';
clearvars -except testscr avginc;
```

3. Create the systematic component of the regression equation

Create the systematic component of the regression equation.

```
y = log(testscr);
N_obs = size(y,1);
x_0 = ones(N_obs,1);
X = [x_0 avginc];
```

4. Estimate the coefficients and the variance-covariance matrix of them

To calculate standard errors of coefficient estimates in nonlinear forms, the Delta method is used. Review the method here `https://www.stata.com/manuals13/rnlcom.pdf`. In particular consider the variance formula at the end of the page. It has two components. The first term is a variance term. It is the variance estimate of $\hat{\beta}_1$. Obtain $\hat{\beta}_1$. Obtain the variance estimate of $\hat{\beta}_1$. The second term is the subject of the next section.

```
LSS = exercisefunction(y,X);
B_hat_x_1 = LSS.B_hat(2,1);
B_hat_x_1_VE = LSS.B_hat_VCE(2,2);
```

5. Obtain the derivative of the nonlinear function of the population coefficient of the independent variable

We need to obtain the derivative of the nonlinear function of the population coefficient of the independent variable. This function is the discrete effect of `x_1`. In the code presented at the end of the section, the first line specifies the amount of the discrete change in `x_1`. What is listed in the remaining of the code is the symbolic calculation of the derivative of the nonlinear function with respect to the parameter of interest, and the evaluation of this derivative at the estimate of this parameter.

```
DC_in_x_1 = 5;
B_x_1 = sym('B_x_1');
DE_of_x_1 = exp(B_x_1*DC_in_x_1)-1;
G = DE_of_x_1;
diff(G)
diff_G_hat = DC_in_x_1*exp(B_hat_x_1*DC_in_x_1);
```

6. Apply the Delta method

Given the two terms required to calculate the variance of the nonlinear function of interest, we are ready to apply the delta method. After calculating this variance, calculate the standard error.

```
DE_x_1_VE = diff_G_hat*B_hat_x_1_VE*diff_G_hat;
DE_x_1_SEE = sqrt(DE_x_1_VE);
```

7. Hypothesis test on the nonlinear function of the population coefficient of the independent variable

Carry out the t test on the discrete effect of `x_1`. What do you conclude?

```
DE_of_x_1_hat = exp(B_hat_x_1*DC_in_x_1)-1;
t = DE_of_x_1/DE_of_x_1_SEE;
t_df = N_obs-size(X,2);
t_critical = tinv(0.975,t_df);
p = tcdf(t,t_df,'upper')*2;
p_critical = 0.05;
```