SOLUTIONS TO EXERCISE 1

FENG LI

SEPTEMBER 10, 2013

2.7

The term "linear" regression always means a regression that is linear in the parameters(details in p. 42).

a. This is a linear regression model. We will see this more clearly if we take a nature log (log to the base e), we have

$$\ln(Y_i) = \ln\left(e^{\beta_1+\beta_2 X_1+u_i}\right) = \beta_1 + \beta_2 X_i + u_i$$

b. We make a little transformation

$$\frac{1-Y_i}{Y_i} = e^{\beta_1+\beta_2 X_i+u_i}$$

and take a nature log, we get the linear model

$$\log\left(\frac{1-Y_i}{Y_i}\right) = \beta_1 + \beta_2 X_i + u_i$$

The two-step is also called *logit transformation*.

c. This is a linear transformation since to the parameters(β_1 and β_2) it is a linear form.

d. This is a nonlinear regression model.

e. This is also nonlinear regression model because β_2 is raised to the third power.

2.13

It is a sample regression line because it is based on a sample of 46 years of observations. The scatter points around the regression line are the actual data points. The difference between the actual consumption expenditure and that estimated from the regression line represents the (sample) residual. Besides GDP, factors such as wealth, interest rate, etc. might also affect consumption expenditure.

3.1

Consider the linear model $Y_i = \beta_1 + \beta_2 X_i + u_i$,

---

School of Statistics and Mathematics, Central University of Finance and Economics. E-mail: m@feng.li.

a. given $E(u_i|X_i) = 0$, we have

$$\begin{aligned} E(Y_i|X_i) &= E\left(\left(\beta_1 + \beta_2 X_i + u_i\right)|X_i\right) \\ &= E(\beta_1|X_i) + E(\beta_2 X_i|X_i) + E(u_i|X_i) \\ &= E(\beta_1) + E(\beta_2 X_i) + E(u_i|X_i) \\ &= \beta_1 + \beta_2 X_i \end{aligned}$$

This is because β_1 and β_2 are constants and X_i is non-stochastic.

b. given $cov(u_i, u_j) = 0$ for all $i \neq j$, which means u_i, u_j are uncorrelated, we have

$$\begin{aligned} \operatorname{cov}(Y_i, Y_j) &= E\left((Y_i - E(Y_i))(Y_j - E(Y_j))\right) \\ &= E\left((Y_i - (\beta_1 + \beta_2 X_i))(Y_j - (\beta_1 + \beta_2 X_j))\right) && \text{by (a)} \\ &= E(u_i u_j) \\ &= E(u_i) E(u_i) && u_i, u_j \text{ uncorrelated} \\ &= 0 && E(u_i) = 0 \end{aligned}$$

c. by assumptions we can get that

$$\begin{aligned} \operatorname{var}(Y_i|X_i) &= E\left((Y_i - E(Y_i))^2\right) \\ &= E\left(u_i^2\right) = \operatorname{var}(u_i) + (Eu_i)^2 \\ &= \sigma^2 && var(u_i) = \sigma^2 \text{ and } E(u_i) = 0 \end{aligned}$$

3.6

By E.q. (3.1.6), for the model $Y_i = \alpha + \beta_{yx} X_i + u_i$, we have

$$\hat{\beta}_{yx} = \frac{\sum x_i y_i}{\sum x_i^2}$$

where $x_i = X_i - \bar{X}$, $y_i = Y_i - \bar{Y}$ and for the model $X_i = \alpha + \beta_{xy} Y_i + \epsilon_i$, we have

$$\hat{\beta}_{xy} = \frac{\sum x_i y_i}{\sum y_i^2}$$

Multiplying the two, we obtain the expression that

$$\hat{\beta}_{yx}\hat{\beta}_{xy} = \frac{\sum x_i y_i}{\sum x_i^2}\frac{\sum x_i y_i}{\sum y_i^2} = \frac{\left(\sum x_i y_i\right)^2}{\sum x_i^2 \sum y_i^2}$$

which is r^2 in E.q. (3.5.8)

3.7

Even though $\hat{\beta}_{yx}\hat{\beta}_{xy} = 1$. This does not mean $\hat{\beta}_{yx} = \hat{\beta}_{xy} = 1$. It may still matter if Y is regressed on X or X on Y.

3.14

The residuals and fitted values of Y will not change. Consider the regression models $Y_i = \beta_1 + \beta_2 X_i + u_i$ and $Y_i = \alpha_1 + \alpha_2 Z_i + \epsilon_i$, where $Z = 2X$ From E.q. (3.1.6), we know that for the slopes

$$\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2}$$
$$\hat{\alpha}_2 = \frac{\sum z_i y_i}{\sum z_i^2} = \frac{\sum (2x_i) y_i}{\sum (2x_i)^2} = \frac{\sum x_i y_i}{2\sum x_i^2} = \frac{1}{2}\hat{\beta}_2$$

But the intercepts are unchanged, by E.q. (3.1.7)

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}$$
$$\hat{\alpha}_1 = \bar{Y} - \hat{\alpha}_2 \bar{Z} = \bar{Y} - \frac{1}{2}\hat{\beta}_2 \left(2\bar{X}\right), \qquad \bar{Z} = 2\bar{X}$$

NOTE: be careful the notation of the lower case x_i and upper case X_i

3.16

a. FALSE. The covariance can be any value depending on the unit of measurement (e.g A and B). By definition the covariance is

$$\text{cov}(A, B) = E\left(A - E\left(A\right)\right)\left(B - E\left(B\right)\right) = E\left(AB\right) - E\left(A\right)E\left(B\right)$$

while the correlation is restricted to $[-1, 1]$ and

$$cor\left(A, B\right) = \frac{\text{cov}(A, B)}{\sqrt{\text{var}\left(A\right)\text{var}\left(B\right)}} = \frac{E\left(AB\right) - E\left(A\right)E\left(B\right)}{\sqrt{\left(E\left(A^2\right) - \left(E\left(A\right)\right)^2\right)\left(E\left(B^2\right) - \left(E\left(B\right)\right)^2\right)}}$$

b. FALSE. Look at the Figure 3.10h (p. 78) The correlation coefficients is a measure of *linear* relationship between two variable. Given nonlinear relationship $Y = X^2$, we can calculate the correlation is 0 but not means Y and X has no relationship.

c. TRUE. Consider a linear regression model $Y_i = \beta_1 + \beta_2 X_i + u_i$, we obtain the estimator $\hat{\beta}_1$ and $\hat{\beta}_2$ where $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$. Now let's regress Y_i on $\hat{Y}_i$. Say we want to make a linear regression on the model

$$Y_i = \alpha_1 + \alpha_2 \hat{Y}_i + \varepsilon_i$$

To get the estimator $\hat{\alpha}_1$ and $\hat{\alpha}_2$, just do some basic algebra as below

$$\hat{\alpha}_2 = \frac{\sum y_i \hat{y}_i}{\sum \hat{y}_i^2} \qquad \text{by Eq. 3.1.6, p. 58}$$
$$= \frac{\sum y_i \left(\hat{\beta}_2 x_i\right)}{\sum \left(\hat{\beta}_2 x_i\right)^2} = \frac{\hat{\beta}_2 \sum x_i y_i}{\hat{\beta}_2^2 \sum x_i^2}$$
$$= \frac{\sum x_i y_i}{\hat{\beta}_2 \sum x_i^2} = 1 \qquad \text{why?}$$

and by Eq. 3.1.10 (p. 64)

$$\hat{\alpha}_1 = \bar{Y} - \hat{\alpha}_2\bar{\hat{Y}} = \bar{Y} - \bar{\hat{Y}} = 0$$

3.19

a. The slope value of 2.250 suggests that over the period 1985-2005, for every unit increase in the ratio of the US to Canadian CPI, on average, the Canadian to US dollar exchange rate ratio increased by about 2.250 units. That is, as the US dollar strengthened against the Basic Econometrics, Gujarati and Porter 24 Canadian dollar, one could get more Canadian dollars for each US dollar. Literally interpreted, the intercept value of -0.912 means that if the relative price ratio were zero, a US dollar would exchange for - 0.912 Canadian dollars (would lose money). Of course, this interpretation is not economically meaningful. With a fairly low to moderate r^2 of 0.440, we should realize that there is a lot of variability in this result.

b. The positive value of the slope coefficient makes economic sense because if U.S. prices go up faster than Canadian prices, domestic consumers will switch to Canadian goods because they can buy more, thus increasing the demand for GM, which will lead to appreciation of the German mark. This is the essence of the theory of purchasing power parity (PPP), or the law of one price.

c. In this case the slope coefficient is expected to be negative, for the higher the Canadian CPI relative to the U.S. CPI, the lower the relative inflation rate in Canada which will lead to depreciation of the U.S. dollar. Again, this is in the spirit of the PPP.