# SOLUTIONS TO EXERCISE 5

FENG LI

OCTOBER 28, 2013

## 10.10

**a.** No. Multicollinearity refers to linear association among variables. Here the association is nonlinear.

**b.** There is no reason to drop them. They are theoretically as well as statistically significant in the present example.

**c.** If one of the variables is dropped, there will be specification bias that will show up in the coefficient(s) of the remaining variable(s).

## 10.12

**a.** False. If exact linear relationship(s) exist among variables, we cannot even estimate the coefficients or their standard errors.

**b.** False. One may be able to obtain one or more significant t values.

**c.** False. As noted in the chapter (see Eq. 7.5.6), the variance of an OLS estimator is given by the following formula:

$$\operatorname{var}\left(\hat{\beta}_j\right) = \frac{\sigma^2}{\sum x_j^2}\left(\frac{1}{1-R_j^2}\right)$$

As can be seen from this formula, a high $R^2$ can be counterbalanced by a low $\sigma^2$ or high $\sum x_j^2$ .

**d.** Uncertain. If a model has only two regressors, high pairwise correlation coefficients may suggest multicollinearity. If one or more regressors enter non-linearly, the pairwise correlations may give misleading answers.

**e.** Uncertain. If the observed collinearity continues in the future sample values, then there may be no harm. But if that is not the case or if the objective is precise estimation, then multicollinearity may be problem.

**f.** False. See answer to (c) above.

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

**g.** False. VIF and TOL provide the same information.

**h.** False. One usually obtains high $R_2$s in models with highly correlated regressors.

**i.** True. As you can see from the formula given in (c), if the variability in $X_3$ is small, $R_j^2$ will tend to be small and in the extreme case of no variability in $X_3$, $\sum x_{3i}^2$ will be zero, in which case the variance of the estimated $\beta_3$ will be infinite.

10.19

**a.** Since ther third regressor $(M_t - M_{t-1})$ is a linear combination of $M_t$ and $M_{t-1}$, where might be a collinearity problem.

**b.** To avoid this, we can just build a model including $M_t$ and $M_{t-1}$.

**c.** All the parameters can be estimated uniquely, as there is no longer perfect collinearity.

**d.** The answer is same as (c.)

10.21

Since from Eq. 7.4.12 and Eq. 7.4.15 for $j = 2, 3$ we have

$$\operatorname{var}\left(\hat{\beta}_j\right) = \frac{\sigma^2}{\sum x_{ji}^2 \left(1 - r_{23}^2\right)}$$

when perfect collinearity exists, $r_{23} \to 1$ which implies

$$\operatorname{var}\left(\hat{\beta}_j\right) \to \infty$$

11.1

**a.** False. The estimators are unbiased but are inefficient.

**b.** True. See Sec. 11.4

**c.** False. Typically, but not always, will the variance be overestimated. See Sec. 11.4 and Exercise 11.9

**d.** False. Besides heteroscedasticity, such a pattern may result from autocorrelation, model specification errors, etc.

**e.** True. Since the true $\sigma_i^2$ are not directly observable, some assumption about the nature of heteroscedasticity is inevitable.

**f.** True. See answer to (d) above.

**g.** False. Heteroscedasticity is about the variance of the error term ui and not about the variance of a regressor.

11.2

**a.** As equation (1) shows, as N increases by a unit, on average, wages increase by about 0.009 dollars. If you multiply the second equation through by N, you will see that the results are quite similar to Eq. (1).

**b.** Apparently, the author was concerned about heteroscedasticity, since he divided the original equation by N. This amounts to assuming that the error variance is proportional to the square of N. Thus the author is using weighted least-squares in estimating Eq. (2).

**c.** The intercept coefficient in Eq. (1) is the slope coefficient in Eq. (2) and the slope coefficient in Eq. (1) is the intercept in Eq. (2).

**d.** No. The dependent variables in the two models are not the same.

11.6

**a.** The assumption made is that the error variance is proportional to the square of GNP, as is described in the postulation. The authors make this assumption by looking at the data over time and observing this relationship.

**b.** The results are essentially the same, although the standard errors for two of the coefficients are lower in the second model; this may be taken as empirical justification of the transformation for heteroscedasticity.

**c.** No. The $R^2$ terms may not be directly compared, as the dependent variables in the two models are not the same.

11.21

The calculated test statistic,

$$\lambda = F = \frac{RSS_2/df}{RSS_1/df} = \frac{140/25}{55/25} = 2.5454$$

The 5% critical F for 25 d.f. in the numerator and denominator is 1.97. Since the estimated value of 2.5454 exceeds this critical value, reject the null of homoscedasticity.

12.2

For $n = 50$ and $k' = 4$, and $\alpha = 5\%$, the critical d values are:

$$d_L = 1.384, \; 4 - d_L = 2.62$$
$$d_U = 1.724, \; 4 - d_U = 2.28$$

(a) positive autocorrelation; (b) inconclusive, (c) inconclusive; and (d) negative autocorrelation.

12.2

**a.** There is serial correlation in Model A, but not in Model B.

**b.** The autocorrelation may be due to misspecification of Model A because it excludes the quadratic trend term.

**c.** One would need prior knowledge of the probable functional form.

12.11

**a.** The figure shows that there is probably specification bias due to a misspecification of the functional form.

**b.** Introduce [log(output)] as an additional regressor. This probably will pick up the quadratic nature of the relationship between cost and output.

## 12.12

**a.** There are many reasons for an outlier. It may be an observation that is simply very different from the rest of the sample; it may be the result of measurement error, or it may be due to poor sampling.

**b.** The observation should not be discarded unless there is some plausible reason to believe that it is erroneous (e.g., measured incorrectly, recorded in error, etc).

**c.** No. The outlier may dominate the RSS.