# Introduction to Econometrics (4th Edition)

by

James H. Stock and Mark W. Watson

# Solutions to Odd-Numbered End-of-Chapter Exercises: Chapter 6

(This version September 14, 2018)

6.1. By equation (6.15) in the text, we know

$$\bar{R}^2 = 1 - \frac{n-1}{n-k-1}(1-R^2).$$

Thus, the values of $\bar{R}^2$ for columns (1)–(3) are: 0.1648, 0.1817, 0.1843.

6.3. (a) On average, a worker earns \$0.61/hour more for each year he ages.

(b) Sally's earnings prediction is $0.11 + 10.44\times1 - 4.56\times1 + 0.61\times29 = 23.68$ dollars per hour.

Betsy's earnings prediction is $0.11 + 10.44\times1 - 4.56\times1 + 0.61\times34 = 26.73$ dollars per hour. The difference is 3.05 \$/hour ($= 0.61\times(34-29)$).

6.5. (a) \$23,400 (recall that *Price* is measured in \$1000s).

(b) In this case $\Delta BDR = 1$ and $\Delta Hsize = 100$. The resulting expected change in price is $23.4 + 0.156 \times 100 = 39.0$ thousand dollars or \$39,000.

(c) The loss is \$48,800.

(d) From the text $\bar{R}^2 = 1 - \frac{n-1}{n-k-1}(1-R^2)$, so $R^2 = 1 - \frac{n-k-1}{n-1}(1-\bar{R}^2)$, thus $R^2 = 0.727$.

6.7. (a) The proposed research in assessing the presence of sex bias in setting wages is too limited. There might be some potentially important determinants of salaries: type of engineer, amount of work experience of the employee, and education level. The sex with the lower wages could reflect the type of engineer, the amount of work experience of the employee, or the education level of the employee. The research plan could be improved with the collection of additional data as indicated above; an appropriate statistical technique for analyzing the data would then be a multiple regression in which the dependent variable is wages and the independent variables would include a binary variable for sex, indicator variables for type of engineer, work experience and education level (highest grade level completed, for example). The potential importance of the suggested omitted variables makes a "difference in means" test inappropriate for assessing the presence of sex bias in setting wages.

(b) The description suggests that the research goes a long way towards controlling for potential omitted variable bias. Yet, there still may be problems. Omitted from the analysis are characteristics associated with behavior that led to incarceration (excessive drug or alcohol use, gang activity, and so forth) that might be correlated with future earnings. Ideally, data on these variables should be included in the analysis as additional control variables.

6.9. For omitted variable bias to occur, two conditions must be true: $X_1$ (the included regressor) is correlated with the omitted variable, and the omitted variable is a determinant of the dependent variable. Since $X_1$ and $X_2$ are uncorrelated, the OLS estimator of $\beta_1$ does not suffer from omitted variable bias.

6.11. (a) $\sum (Y_i - b_1 X_{1i} - b_2 X_{2i})^2$

(b)

$$\frac{\partial \sum (Y_i - b_1 X_{1i} - b_2 X_{2i})^2}{\partial b_1} = -2\sum X_{1i}(Y_i - b_1 X_{1i} - b_2 X_{2i})$$
$$\frac{\partial \sum (Y_i - b_1 X_{1i} - b_2 X_{2i})^2}{\partial b_2} = -2\sum X_{2i}(Y_i - b_1 X_{1i} - b_2 X_{2i})$$

(c) From (b), $\hat{\beta}_1$ satisfies $\sum X_{1i}(Y_i - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i}) = 0$ , or $\hat{\beta}_1 = \frac{\sum X_{1i}Y_i - \hat{\beta}_2 \sum X_{1i}X_{2i}}{\sum X_{1i}^2}$

and the result follows immediately.

(d) Following analysis as in (c) $\hat{\beta}_2 = \frac{\sum X_{2i}Y_i - \hat{\beta}_1 \sum X_{1i}X_{2i}}{\sum X_{2i}^2}$ and substituting this into the

expression for $\hat{\beta}_1$ in (c) yields $\hat{\beta}_1 = \frac{\sum X_{1i}Y \frac{\sum X_{2i}Y_i - \hat{\beta}_1 \sum X_{1i}X_{2i}}{\sum X_{2i}^2} \sum X_{1i}X_{2i}}{\sum X_{1i}^2}$.

Solving for $\hat{\beta}_1$ yields: $\hat{\beta}_1 = \frac{\sum X_{2i}^2 \sum X_{1i}Y_i - \sum X_{1i}X_{2i} \sum X_{2i}Y_i}{\sum X_{1i}^2 \sum X_{2i}^2 - (\sum X_{1i}X_{2i})^2}$

(continued on the next page)

6.11 (continued)

(e) The least squares objective function is $\sum (Y_i - b_0 - b_1X_{1i} - b_2X_{2i})^2$ and the partial derivative with respect to $b_0$ is

$$\frac{\partial \Sigma (Y_i - b_0 - b_1X_{1i} - b_2X_{2i})^2}{\partial b_0} = -2\sum (Y_i - b_0 - b_1X_{1i} - b_2X_{2i}).$$

Setting this to zero and solving for $\hat{\beta}_0$ yields: $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X}_1 - \hat{\beta}_2\bar{X}_2$.

(f) Substituting $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X}_1 - \hat{\beta}_2\bar{X}_2$.into the least squares objective function yields $\sum (Y_i - \hat{\beta}_0 - b_1X_{1i} - b_2X_{2i})^2 = \sum \left((Y_i - \bar{Y}) - b_1(X_{1i} - \bar{X}_1) - b_2(X_{2i} - \bar{X}_2)\right)^2$, which is identical to the least squares objective function in part (a), except that all variables have been replaced with deviations from sample means. The result then follows as in (c).

Notice that the estimator for $\beta_1$ is identical to the OLS estimator from the regression of $Y$ onto $X_1$, omitting $X_2$. Said differently, when $\sum (X_{1i} - \bar{X}_1)(X_{2i} - \bar{X}_2) = 0$, the estimated coefficient on $X_1$ in the OLS regression of $Y$ onto both $X_1$ and $X_2$ is the same as estimated coefficient in the OLS regression of $Y$ onto $X_1$.