CHAPTER 12

Understanding and Interpreting Regression with Two x's

12.0 What We Need to Know When We Finish This Chapter 453

12.1 Introduction 456

12.2 The Variances of b_1 and b_2 456

12.3 The Interaction between x_{1i} and x_{2i} 463

12.4 Estimated Standard Deviations 468

12.5 Unrestricted and Restricted Regressions 470

12.6 Joint Hypothesis Tests 475

12.7 Wait! What If It's All a Mistake? 484

12.8 What Happens to Chapters 8, 9, and 10 Now? 486

12.9 Conclusion 491

Exercises 492

12.0 What We Need to Know When We Finish This Chapter

The slopes that we obtain when we minimize the sum of squared errors for the regression of equation (11.12) are best linear unbiased (BLU) estimates of the population coefficients. If the population relationship includes two explanatory variables, the precision of these slopes depends heavily on the extent to which the two explanatory variables are related. Including an irrelevant variable is inefficient, but does not create bias. Everything that we have done in chapters 8 through 10 holds with two explanatory variables, either exactly or with minor, sensible extensions. Here are the essentials.

1. **Section 12.3:** If x_{1i} and x_{2i} are highly correlated, it's often called *multicollinearity*. If we omit either, we bias the estimated effect of the other. If we include both, their estimated effects are unbiased but may have large variances, especially if n is small. In general, the only responsible way to achieve greater precision is to increase n. Multicollinearity cannot be responsible for slopes with implausible signs or magnitudes and cannot create spurious significance.
2. **Equation (12.15), section 12.4:** The sample estimate of σ^2 is

$$s^2 = \frac{\sum_{i=1}^{n}\left(y_i - \left(a + b_1 x_{1i} + b_2 x_{2i}\right)\right)^2}{n-3} = \frac{\sum_{i=1}^{n} e_i^2}{n-3}.$$

3. **Equation (12.16), section 12.4:** The sample standard deviation of b_1 is

$$\mathrm{SD}\left(b_1\right) = +\sqrt{\frac{s^2}{\sum_{i=1}^{n}\left(x_{1i}-\bar{x}_1\right)^2 - \frac{\left(\sum_{i=1}^{n}\left(x_{2i}-\bar{x}_2\right)\left(x_{1i}-\bar{x}_1\right)\right)^2}{\sum_{i=1}^{n}\left(x_{2i}-\bar{x}_2\right)^2}}}$$

$$= +\sqrt{\frac{s^2}{\sum_{i=1}^{n}\left(e_{(x_1 x_2)i} - \bar{e}_{(x_1 x_2)}\right)^2}}.$$

4. **Section 12.5:** *Joint hypotheses* specify values for two or more parameters simultaneously.
5. **Equation (12.23), section 12.5:** A restricted regression adopts a null hypothesis regarding the value or values of one or more parameters. This null hypothesis may also be referred to as an assumption or, most commonly, a *restriction*. The sum of squared errors from a restricted regression is always at least as large as the sum of squared errors from

an unrestricted regression:

$$\left(\sum_{i=1}^{n} e_i^2\right)_R \geq \left(\sum_{i=1}^{n} e_i^2\right)_U.$$

In particular, the sum of squared errors, and therefore R^2, can never go down when another explanatory variable is added to the regression. However, the adjusted R^2 can go down.

6. **Equation (12.29), section 12.6:** The test of j restrictions is

$$\frac{\left(\dfrac{\left(\sum_{i=1}^{n} e_i^2\right)_R - \left(\sum_{i=1}^{n} e_i^2\right)_U}{j}\right)}{\left(\dfrac{\left(\sum_{i=1}^{n} e_i^2\right)_U}{n-3}\right)} \sim F(j, n-3).$$

7. **Section 12.7:** If we include an irrelevant variable in our regression, the slopes are still unbiased estimators of the true coefficient values. In particular, the slope for the irrelevant variable should be pretty close to zero, at least statistically. However, the inclusion of an irrelevant variable will usually reduce the precision of the estimated effects of relevant variables.

8. **Section 12.8:** With two explanatory variables, ordinary least squares (OLS) slopes are still unbiased estimators if the disturbances are heteroscedastic or autocorrelated. The White test, the White heteroscedasticity-consistent variance estimator, and the Newey-West autocorrelation-consistent variance estimator are all still valid, but need to be reformulated to incorporate the second explanatory variable. Weighted least squares (WLS) or generalized least squares (GLS) are still required to obtain best linear unbiased estimators.

9. **Equation (12.47), section 12.8:** When one explanatory variable is endogenous, the other explanatory variable must be included in the instrumenting equation, along with the instrumental variable itself.

10. **Equations (12.50) and (12.51), section 12.8:** If both explanatory variables are endogenous, we need at least two instrumental variables.