# Problem Set 14
# "Multiple Regression: More Tests and Functional Forms"

(14.1) [Quadratic functional form] Suppose we have estimated the following equation for average cost as a function of output, using a sample of 20 firms. Point estimates (and standard errors in parentheses) are given below. Here, AC denotes average cost and Q denotes output quantity. We assume the error term is normally-distributed.

$$AC = \underset{(0.42)}{8.25} - \underset{(0.05)}{0.50\,Q} + \underset{(0.004)}{0.010\,Q^2}$$

a. Test the null hypothesis that average cost is a linear function of output against the alternative that it is a quadratic function, at 5% significance. Is this a one-tailed test or a two-tailed test? Give
   - the *value* of the test statistic
   - the *degrees of freedom*
   - the *critical point(s)* from the appropriate table at the back of your textbook (or compute the *p-value* using a spreadsheet program)
   - your conclusion: whether you reject the null hypothesis at 5% significance.

b. At what level of output Q* does average cost reach its minimum, according to these estimates? What is the minimum level of average cost AC* ?

(14.2) [Testing hypotheses about more than one coefficient] Suppose we wish to estimate the demand for health care using the following equation.

$$\ln(qhealth) = \beta_1 + \beta_2 \ln(phealth) + \beta_3 \ln(pother) + \beta_4 \ln(inc)$$

Here, *qhealth* denotes the quantity of health care, *phealth* denotes the price of health care, *pother* denotes the price of other goods, and *inc* denotes income. Now the theory of consumer demand shows that if income and all prices double, then the quantity demanded of any good will remain unchanged (because the consumer's budge set is unchanged).

a. Prove that demand theory implies the following restriction on the coefficients: $\beta_2 + \beta_3 + \beta_4 = 0$. [Hint: Recall that ln(2x) = ln(2) + ln(x). So first substitute (ln(2)+ln(phealth)) for ln(phealth), substitute (ln(2)+ln(pother)) for ln(pother) and substitute (ln(2)+ln(inc)) for ln(inc). Next show that if the right-hand side still equals ln(qhealth), then it follows that $\beta_2 + \beta_3 + \beta_4 = 0$.]

To test this restriction, we need an estimate of $(\beta_2 + \beta_3 + \beta_4)$ and its standard error. Let $\theta = (\beta_2 + \beta_3 + \beta_4)$, or equivalently $\beta_2 = (\theta - \beta_3 - \beta_4)$.

b. Substitute $(\theta - \beta_3 - \beta_4)$ for $\beta_2$ in the regression equation, and rearrange so that $\theta$ is one of the coefficients. What are the new regressors in this transformed equation?

(14.3) [Testing hypotheses about more than one coefficient] Suppose we wish to estimate a production function using the following equation:

$$\ln(\text{output}) = \beta_1 + \beta_2 \ln(\text{labor}) + \beta_3 \ln(\text{capital}) + \beta_4 \ln(\text{energy})$$

Now production is characterized by "constant returns to scale" if when doubling inputs causes output to double as well.

a. Prove that "constant returns to scale" implies the following restriction on the coefficients: $\beta_2 + \beta_3 + \beta_4 = 1$. [Hint: Recall that ln(2x) = ln(2) + ln(x). So first substitute (ln(2)+ln(labor)) for ln(labor), substitute (ln(2)+ln(capital)) for ln(capital), and substitute (ln(2)+ln(energy)) for ln(energy). Next show that if the right-hand side equals (ln(2)+ln(output)), then it follows that $\beta_2 + \beta_3 + \beta_4 = 1$.]

To test this restriction, we need both an estimate of $(\beta_2 + \beta_3 + \beta_4 - 1)$ and its standard error. Let $\theta = (\beta_2 + \beta_3 + \beta_4 - 1)$, or equivalently $\beta_2 = (\theta - \beta_3 - \beta_4 + 1)$.

b. Substitute $(\theta - \beta_3 - \beta_4 + 1)$ for $\beta_2$ in the regression equation, and rearrange so that $\theta$ is one of the coefficients. What are the new regressors in this transformed equation? What is the new dependent variable?

(14.4) [Confidence intervals, hypothesis tests, functional forms] Suppose an equation for the demand for telephone calls is estimated, on a sample of 43 states, with the following results:

$$\ln(Q) = \underset{(1.2)}{3.5} - \underset{(0.3)}{0.8}\ \ln(P) + \underset{(0.2)}{1.1}\ \ln(POP)$$

where ln(.) is the natural logarithm function, Q = quantity of telephone calls, P = average price of telephone calls, and POP = state population. Numbers in parentheses are standard errors of the coefficient estimates. Assume the error term is normally-distributed.

a. Give the degrees of freedom for this problem.
b. Compute the estimated price elasticity of demand and a 95% confidence interval for it.
c. Test the null hypothesis that the quantity of telephone calls demanded is proportional to the state population, holding price constant—that is, that the elasticity with respect to population is one—against the alternative hypothesis that it is not proportional, at 5% significance. Is this a one-tailed test or a two-tailed test? Give
   - the *value* of the test statistic
   - the *critical point(s)* from the appropriate table at the back of your textbook (or compute the *p-value* using a spreadsheet program)
   - your conclusion: whether you reject the null hypothesis at 5% significance.
d. Suppose you believe the number of telephone calls demanded is indeed proportional to the state's population, holding price constant. What equation could you estimate that incorporates this restriction? (Define any new variables you use.)

(14.5) [Confidence intervals, hypothesis tests, functional forms]  Suppose an equation for the demand for electric power is estimated, on a sample of 63 cities, with the following results:

$$\ln(Q) = \underset{(0.2)}{2.5} - \underset{(0.3)}{0.7} \ln(P) + \underset{(0.1)}{0.9} \ln(I)$$

where ln(.) is the natural logarithm function, Q = quantity of electricity consumed per capita by residential consumers, P = average price of electricity, and I = income per capita.  Numbers in parentheses are standard errors of the coefficient estimates.  Assume the classical assumptions hold and that the error term is normally-distributed.

a. Give the degrees of freedom for this problem.
b. Compute the estimated price elasticity of demand and a 95% confidence interval for it.
c. Test the null hypothesis that the price elasticity of demand equals zero ("perfectly inelastic demand") against the alternative hypothesis that it is negative, at 5% significance.  Is this a one-tailed test or a two-tailed test?  Give
   - the *value* of the test statistic
   - the *critical point(s)* from the appropriate table at the back of your textbook (or compute the *p-value* using a spreadsheet program)
   - your conclusion:  whether you reject the null hypothesis at 5% significance.
d. Compute the estimated income elasticity of demand and a 95% confidence interval for it.
e. Test the null hypothesis that the income elasticity of demand equals one against the alternative hypothesis that it is not equal to one, at 5% significance.  Is this a one-tailed test or a two-tailed test?  Give
   - the *value* of the test statistic
   - the *critical point(s)* from the appropriate table at the back of your textbook (or compute the *p-value* using a spreadsheet program)
   - your conclusion:  whether you reject the null hypothesis at 5% significance.

(14.6) [Useful functional forms, dummy variables] Suppose we have estimated the demand for gasoline, using a sample of 500 households across the nation, with the following results. Here, *qg* denotes the quantity of gasoline purchased in gallons per month, *pg* denotes the price of gasoline per gallon, *inc* denotes monthly income, and *drural* is a dummy variable equal to one if the household lives in a rural area and zero otherwise. (Standard errors are given in parentheses.)

$$\ln(qg) = \underset{(0.4)}{0.6} - \underset{(0.2)}{0.5}\ln(pg) + \underset{(0.3)}{0.9}\ln(inc) + \underset{(0.16)}{0.2}\,drural$$

a. Can you compute estimates of price elasticity of demand and the income elasticity of demand from the above information alone? If yes, compute the estimates. If no, indicate what other information you need.
b. Test the "Law of Demand" that price and quantity are negatively related, at 5% significance. Is this a one-tailed test or a two-tailed test? Give
   - the *value* of the test statistic
   - the *critical point(s)* from the appropriate table at the back of your textbook (or compute the *p-value* using a spreadsheet program)
   - your conclusion: whether you reject the null hypothesis at 5% significance.
c. Test whether gasoline is a normal good (income and quantity demanded are positively related) at 5% significance. Is this a one-tailed test or a two-tailed test? Give the test statistic, the critical point(s), and the conclusion.
d. Test the hypothesis that rural and urban households have the same demand curve, against the alternative hypothesis that they have different demand curves, at 5% significance. Is this a one-sided test or a two-sided test? Give
   - the *value* of the test statistic
   - the *critical point(s)* from the appropriate table at the back of your textbook (or compute the *p-value* using a spreadsheet program)
   - your conclusion: whether you reject the null hypothesis at 5% significance.

(14.7) [Confidence intervals, hypothesis tests, functional forms, dummy variables, structural change] A wage equation was estimated on a sample of 200 workers with the following results:

$$\ln(W) = \underset{(0.4)}{0.6} + \underset{(0.02)}{0.08\,S} + \underset{(0.03)}{0.05\,X} - \underset{(0.02)}{0.001\,X^2} + \underset{(0.05)}{0.12\,D}$$

where ln(.) is the natural logarithm function, W = wage or hourly earnings, S = years of schooling, X = years of labor-market experience, and D = a dummy variable equal to one if the worker is a union member and zero if the worker is not a union member. Numbers in parentheses are standard errors of the coefficient estimates.

a. By approximately how much do wages increase with one year of additional year of schooling, holding all other variables constant? Compute an estimate of the percent increase and a 95% confidence interval for your answer.
b. By approximately how much do wages increase with one additional year of labor-market experience, for a worker with 10 years of labor-market experience already, holding all other variables constant? Compute an estimate of the percent increase.
c. According to the coefficient estimates, which group earns more: union workers or non-union workers, holding schooling and experience constant?
d. Test the null hypothesis that union and non-union workers with the same schooling and experience earn the same wages, against the alternative hypothesis that they earn different wages, at 5% significance. Is this a one-sided test or a two-sided test? Give
   - the *value* of the test statistic
   - the *critical point(s)* from the appropriate table at the back of your textbook (or compute the *p-value* using a spreadsheet program)
   - your conclusion: whether you reject the null hypothesis at 5% significance.
e. Suppose you suspected that all coefficients including the intercept were different for union and non-union workers. What equation could you estimate to allow for this difference? (Define any new variables you use.) How many restrictions on the β parameters reduce this equation to the original equation reported above?
f. Suppose the sum of squared residuals for the original equation were 414.0 and the sum of squared residuals for the equation described in part (e) were 384.0. Test the null hypothesis that all workers share the same coefficients for schooling and experience (but not the same intercept), against the alternative hypothesis that all coefficients including the intercept were different for union and non-union workers, at 5% significance. Give the value of the test statistic, its degrees of freedom, the critical point or points, and your conclusion.

(14.8) [Useful functional forms, dummy variables] Suppose we have estimated an equation for weekly clothing expenditures using a sample of 325 households in a particular city. Point estimates (and standard errors in parentheses) are given below. Here, E denotes weekly clothing expenditures, I denotes weekly income, N denotes the total number of persons in the household, C denotes a dummy variable equal to one if there are children in the household (and equal to zero otherwise), and A denotes a dummy variable equal to one if there are persons over 65 in the household (and equal to zero otherwise). Because all families live in the same city, you may assume they all face the same prices for clothing.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ln(E) = | 0.6 | + | 0.85 ln(I) | + | 0.88 ln(N) | - | 0.05 C | - 0.12 A |
| | (0.4) | | (0.05) | | (0.04) | | (0.03) | (0.07) |

a. Can you compute an estimate of the income elasticity of demand from the above information alone? If yes, compute the estimate and a 95% confidence interval for it. If no, indicate what other information you need.
b. Test the null hypothesis that clothing expenditures rise in proportion to household size (number of persons in the household) against the null hypothesis that clothing expenditures rise *less* than in proportion to household size, at 5% significance. Is this a one-tailed test or a two-tailed test? Give
   - the *value* of the test statistic
   - the *critical point(s)* from the appropriate table at the back of your textbook (or compute the *p-value* using a spreadsheet program)
   - your conclusion: whether you reject the null hypothesis at 5% significance.
c. Holding constant I, N, and A, does the presence of children raise or lower clothing expenditures, according to these estimates? By how much?
d. The sum of squared residuals from this equation was 160. To test whether C and A are jointly significant, the same equation was estimated after dropping these regressors. The result was a sum of squared residuals equal to 166. Explain intuitively why the sum of squared residuals rose. Test the null hypothesis that C and A both have zero coefficients, against the alternative hypothesis that at least one has a nonzero coefficient, at 5% significance. Give the test statistic, the critical point(s), and the conclusion.
e. Another researcher tries to replicate these results using a new sample from a different city. However, in the new sample, all households either have children or have a person over 65, but not both. Will this pose a problem for estimation? Explain your answer.

(14.9) [Dummy variables and structural change] Suppose we wish to test the relationship between economic growth rates and tax rates using a sample of 50 U.S. states and 10 Canadian provinces—a total of 60 observations. Let

$y_i$ = recent economic growth rate

$x_i$ = tax rate.

$c_i$ = 1 if observation is a Canadian province, 0 if a U.S. state.

The following equations have been estimated, yielding the sums of squared residuals (SSR) shown at right.

| | | |
|---|---|---|
| (i) | $y_i = \beta_1 + \beta_2 x_i$ | SSR = 144 |
| (ii) | $y_i = \beta_1 + \beta_2 x_i + \beta_3 c_i$ | SSR = 114 |
| (iii) | $y_i = \beta_1 + \beta_2 x_i + \beta_3 (c_i x_i)$ | SSR = 120 |
| (iv) | $y_i = \beta_1 + \beta_2 x_i + \beta_3 c_i + \beta_4 (c_i x_i)$ | SSR = 112 |

a. Assume that U.S. states and Canadian provinces have the same slope. Test the null hypothesis that states and provinces have the same intercept, against the alternative hypothesis that they have different intercepts, at 5% significance.
   - Which model above represents the null hypothesis?
   - Which model above represents the alternative hypothesis?
   - Give the degrees of freedom in the numerator and the denominator.
   - Compute the value of the relevant F statistic.
   - Find the *critical point* from the appropriate table at the back of your textbook (or compute the *p-value* using a spreadsheet program).
   - Can you reject the null hypothesis at 5% significance?

b. Assume that U.S. states and Canadian provinces may have different intercepts. Test the null hypothesis that states and provinces have the same slope, against the alternative hypothesis that they have different slopes, at 5% significance.
   - Which model above represents the null hypothesis?
   - Which model above represents the alternative hypothesis?
   - Give the degrees of freedom in the numerator and the denominator.
   - Compute the value of the relevant F statistic.
   - Find the *critical point* from the appropriate table at the back of your textbook (or compute the *p-value* using a spreadsheet program).
   - Can you reject the null hypothesis at 5% significance?

c. Test the null hypothesis that states and provinces have the same intercept and same slope, against the alternative hypothesis that they have different intercepts and slopes, at 5% significance.
   - Which model above represents the null hypothesis?
   - Which model above represents the alternative hypothesis?
   - Give the degrees of freedom in the numerator and the denominator.
   - Compute the value of the relevant F statistic.
   - Find the *critical point* from the appropriate table at the back of your textbook (or compute the *p-value* using a spreadsheet program).
   - Can you reject the null hypothesis at 5% significance?

(14.10) [Dummy variables and structural change] Suppose we wish to test the relationship between grade-point average and salary after graduation using a sample of 44 recent graduates from Colleges A and B. Let

$salary_i$ = salary after graduation
$gpa_i$ = grade-point average.
$d_i$ = 1 if person graduated from College A, 0 if graduated from College B.

The following equations have been estimated, yielding the sums of squared residuals (SSR) shown at right.

| | | |
|---|---|---|
| (i) | $salary_i = \beta_1 + \beta_2\ gpa_i$ | SSR = 150 |
| (ii) | $salary_i = \beta_1 + \beta_2\ gpa_i + \beta_3\ d_i$ | SSR = 123 |
| (iii) | $salary_i = \beta_1 + \beta_2\ gpa_i + \beta_3\ (d_i\ gpa_i)$ | SSR = 125 |
| (iv) | $salary_i = \beta_1 + \beta_2\ gpa_i + \beta_3\ d_i + \beta_4\ (d_i\ gpa_i)$ | SSR = 120 |

a. Assume that all people in the sample have the same slope. Test the null hypothesis that all people have the same intercept, against the alternative hypothesis that they have different intercepts by college, at 5% significance.
   - Which model above represents the null hypothesis?
   - Which model above represents the alternative hypothesis?
   - Give the degrees of freedom in the numerator and the denominator.
   - Compute the value of the relevant F statistic.
   - Find the *critical point* from the appropriate table at the back of your textbook (or compute the *p-value* using a spreadsheet program).
   - Can you reject the null hypothesis at 5% significance?

b. Assume that people may have different intercepts by college. Test the null hypothesis that all people have the same slope, against the alternative hypothesis that they have different slopes by college, at 5% significance.
   - Which model above represents the null hypothesis?
   - Which model above represents the alternative hypothesis?
   - Give the degrees of freedom in the numerator and the denominator.
   - Compute the value of the relevant F statistic.
   - Find the *critical point* from the appropriate table at the back of your textbook (or compute the *p-value* using a spreadsheet program).
   - Can you reject the null hypothesis at 5% significance?

c. Test the null hypothesis that all people have the same intercept and same slope, against the alternative hypothesis that they have different intercepts and slopes by college, at 5% significance.
   - Which model above represents the null hypothesis?
   - Which model above represents the alternative hypothesis?
   - Give the degrees of freedom in the numerator and the denominator.
   - Compute the value of the relevant F statistic.
   - Find the *critical point* from the appropriate table at the back of your textbook (or compute the *p-value* using a spreadsheet program).
   - Can you reject the null hypothesis at 5% significance?

(14.11) [Omitted variable bias] Suppose the true population regression function relating earnings to education and ability is given by the equation

(i) ln(earnings) = 6.4 + 0.09 education + 0.03 ability + error term #1

and that education and ability are related by the following equation

(ii) ability = 63 + 2.5 education + error term #2

Now suppose the earnings equation (i) is estimated by least squares, but the “ability” variable is omitted for lack of data.

a. Are the least-squares estimators of the intercept and the coefficient of “educ” biased or unbiased in this case?
b. Substitute equation (ii) into equation (i) and simplify.
c. Compute the expected value of the least-squares estimator of the intercept.
d. Compute the expected value of the least-squares estimator of the coefficient of “educ.”

[end of problem set]