(a) A restricted form of the model that yields a direct estimate of β_2 may be obtained by substituting in $\beta_3 = -\beta_2$.

$$Y_i = \beta_1 + \beta_2(D_{2i} - D_{3i}) + \varepsilon_i$$

(b) Denoting the explanatory variable of this restricted model as $X_i = D_{2i} - D_{3i}$, it has the following form.

$$X_i = \begin{cases} -1 & \text{if output is from process } A \\ 1 & \text{otherwise} \end{cases}$$

(c) The correspondence between the two sets of coefficients is exactly identifying:

$$E(Y_i) = \begin{cases} \alpha & = \beta_1 - \beta_2 \quad \text{if output is from process A} \\ \alpha + \beta & = \beta_1 + \beta_2 \quad \text{if output is from process B.} \end{cases}$$

10.33 The correspondence between the two sets of parameters is

$$\begin{aligned} \beta_1 &= \mu + \alpha_1 \\ \beta_1 + \beta_2 &= \mu + \alpha_2 \\ \beta_1 + \beta_3 &= \mu + \alpha_3 \end{aligned}$$

(a) The parameters of the one-way analysis of variance are underidentified. (Given values for the 3 parameters β_1, β_2, and β_3, it is not possible to deduce unique implied values for the 4 parameters μ, α_1, α_2, and α_3.)

(b) In their specification of the one-way analysis of variance model statisticians add the requirement that $\sum \alpha_g = 0$ because this additional restriction serves to identify the 4 parameters.

Examination Questions

Many of the following questions bring together concepts from Chapters 7–10. Some assume the provision of the relevant computer output.

1. You have replicated Hendrick Houthakker's estimation of a loglinear model for electricity demand:

$$\log q_i = \beta_1 + \beta_2 \log m_i + \beta_3 \log p_{1i} + \beta_4 \log p_{2i} + \beta_5 \log h_i + \varepsilon_i.$$

Your estimation of this regression is reproduced below. Also shown is a second regression of the form

$$\log q_i = b_1 + b_2 \log m_i + e_i.$$

You can think of this as an Engle curve that you might have estimated had you been unaware of the variation in the other variables across the towns in Houthakker's data set.

(a) From the multiple regression, what is the estimated income elasticity of demand for electricity? Considering it purely as a point estimate, does it indicate that electricity demand is income elastic or inelastic?

(b) Turning to the simple regression estimation of the Engle curve, what is the estimated income elasticity of demand for electricity?

i. Considering it purely as a point estimate, does it indicate that electricity demand is income elastic or inelastic?

ii. Is the evidence in this estimate strong enough to allow you to reject the hypothesis that electricity is income elastic? State all components of your test procedure carefully.

(c) Do the data support restricting the demand function to its Engle curve form? State all components of your test procedure carefully.

(d) Which of the two estimated income elasticities are you more willing to believe, and why?

2. Consider the Cobb-Douglas production function

$$Q = \gamma K^{\beta} L^{\alpha}. \qquad (*)$$

(a) What parameter restriction corresponds to the hypothesis of constant returns to scale? Would you describe this as a linear restriction?

(b) Show that, using this restriction, it is possible to derive a "labor-intensive" version of the production function. Show how this labor-intensive function can be converted to the form of a linear regression.

(c) Returning to the original unrestricted production function $(*)$,

i. show how it can be converted to the form of a multiple regression in which α and β are slope coefficients.

ii. Show how to rewrite this regression to incorporate the constant-returns-to-scale restriction, in such a way that α is eliminated and a direct estimate of β would be obtained.

iii. Indicate any relationship between this restricted regression and the labor-intensive regression of part (b).

3. In a study of returns to education, a researcher estimates log-lin regressions of the form

$$\log Y_i = \alpha + \beta X_i + \varepsilon_i,$$

where Y is weekly earnings and X is years of schooling.

The researcher begins by estimating the regression using the complete sample of $n = 1000$ people:

$$\text{full sample:} \qquad \log Y_i = \underset{(0.906)}{7.60} + \underset{(0.00110)}{0.075}\, X_i + e_i \qquad \text{SSE} = 2.175.$$

She then estimates separate subsample regressions for males and females, based on $n_1 = 600$ and $n_2 = 400$ observations respectively:

$$\text{males:} \qquad Y_i = \underset{(0.944)}{7.26} + \underset{(0.00102)}{0.072}\, X_i + e_i \qquad \text{SSE} = 0.922$$

$$\text{females:} \qquad Y_i = \underset{(0.891)}{7.02} + \underset{(0.00104)}{0.076}\, X_i + e_i \qquad \text{SSE} = 1.013.$$

Answer the following questions.

(a) According to the coefficient point estimates, which group earns the highest return to education, males or females? Cite the estimates that are the basis for your conclusion.

(b) Perform a test of whether the data support imposing a common model on both males and females. State carefully all parts of your test procedure.

(c) An alternative means of distinguishing between males and females would be to define a dummy variable like

$$D_i = \begin{cases} 0 & \text{if male} \\ 1 & \text{if female} \end{cases}$$

and introduce this into the full-sample regression as follows:

dummy variable regression: $\log Y_i = \alpha + \beta X_i + \alpha^* D_i + \beta^* D_i X_i + \varepsilon_i.$

Under this approach,

i. what would be the least squares estimates of these coefficients?

ii. What would be the SSE of this regression?

(d) In light of the various alternative models we have discussed, describe how you would test the following hypotheses. In each case indicate which of the above regressions is the best to use, state the null and alternative hypotheses in terms of restrictions on the parameters of that regression, and indicate the appropriate test statistic and its distribution, including numerical values for degrees of freedom. (It is not necessary to perform the test.)

i. Males and females earn the same rate of return to education.

ii. Males and females earn the same base income, controlling for education.

4. Beginning with the production function

$$Q = \gamma K^\alpha L^{1-\alpha},$$

the Solow-Swan model of economic growth leads to the following equation explaining the logarithm of living standards q:

$$\log q = \frac{\log \gamma}{1-\alpha} + \frac{\alpha}{1-\alpha} \log s - \frac{\alpha}{1-\alpha} \log(\delta + n).$$

Here $q = Q/L$ is output per worker, s is the economy's savings rate, and n is the population growth rate. With data observed across countries on these variables, and assuming some value for the depreciation rate δ, the equation can be estimated as a multiple regression:

$$\log q_i = \beta_1 + \beta_2 \log s_i + \beta_3 \log(\delta + n_i) + \varepsilon_i.$$

(a) What is the economic interpretation of the parameter α?

(b) Comparing the last two equations, what is the nature of the relationship between the regression coefficients β_1, β_2, β_3 and the parameters α and γ? Specifically,

  i. is the relationship over-, exactly-, or underidentifying?

  ii. What restriction on the regression coefficients β_k is implied?

(c) Suppose the regression is estimated in unrestricted form. Are the implied estimates of α and γ unique?

(d) Show one way that the regression model can be rewritten to incorporate the restriction.

(e) The attached regression output includes estimation results for both the restricted and unrestricted regressions based on a sample of 98 countries.

  i. From the restricted regression,

    A. what are the estimates of β_1, β_2, and β_3, direct or implied?

    B. What is the implied estimate of α? Is it plausible? Explain why or why not.

  ii. Perform a test of the restriction. State all steps in your test procedure.

5. Beginning with the production function

$$Q = \gamma K^{\alpha} H^{\beta} L^{1-\alpha-\beta},$$

Mankiw, Romer, and Weil derived an "augmented" version of the Solow-Swan growth model that yields the following equation explaining the logarithm of living standards q:

$$\log q = \frac{\log \gamma}{1-\alpha-\beta} + \frac{\alpha}{1-\alpha-\beta} \log s_k - \frac{\alpha+\beta}{1-\alpha-\beta} \log(\delta+n) + \frac{\beta}{1-\alpha-\beta} \log s_h.$$

Here $q = Q/L$ is output per worker, s_k is the economy's savings rate for physical capital, n is the population growth rate, and s_h is the savings rate for human capital. With data observed across countries on these variables, and assuming some value for the depreciation rate δ, the equation can be estimated as a multiple regression:

$$\log q_i = \beta_1 + \beta_2 \log s_{ki} + \beta_3 \log(\delta + n_i) + \beta_4 \log s_{hi} + \varepsilon_i.$$

(a) Suppose the regression is estimated in unrestricted form.

  i. Are the implied estimates of α, β, and γ unique?

  ii. Mankiw, Romer, and Weil report the unrestricted estimates $\hat{\beta}_1 = 7.781$, $\hat{\beta}_2 = 0.70$, $\hat{\beta}_3 = -1.50$, and an SSE of 14.680. (Of course they also reported an estimate of β_4, but we will not use this in this question.) Use these estimates to obtain implied estimates of α, β, and γ.

(b) The relationship between the above two equations implies a restriction on the slope coefficients of the regression.

  i. What is this implied restriction on the β_k? Is it linear?

ii. The regression output on the next page reports the estimation of the above regression model with the restriction imposed. Show how the above regression model can be rewritten to incorporate the restriction, in the form estimated below.

iii. Use the restricted estimation results to obtain implied estimates of α, β, and γ.

iv. Perform a test of whether the restriction is supported by the data. State carefully all parts of your test procedure.

6. Consider a demand function of the form

$$\log q_i = \beta_1 + \beta_2 \log p_{1i} + \beta_3 \log p_{2i} + \beta_4 \log m_i + \varepsilon_i,$$

where q_i is quantity of the good purchased, p_{1i} is its price, p_{2i} is the price of some complementary or substitutable commodity, and m_i is income; i indexes households.

(a) What parameter restriction would correspond to each of the following null hypotheses?

i. "The demand curve is downward sloping."

ii. The demand for q_i is price elastic.

iii. Good q_i is a substitute for the good whose price is denoted by p_{2i}.

iv. Good q_i is income inelastic.

(b) Suppose that prices vary across households, but that data on prices are unavailable. Consequently the researcher is limited to estimating an Engle curve by running a regression of q_i on m_i. What are the implications for estimation and hypothesis testing with respect to β_4?

(c) Suppose, by contrast, that all households face the same prices, and that these common values p_1 and p_2 are known. What are the implications for the estimation of β_2, β_3, and β_4?

7. The attached computer output includes results for the following versions of one of the ACMS marginal productivity conditions:

$$\log(V/L)_i = \alpha + \beta \log w_i + \varepsilon_i \tag{A}$$

$$\log(V/L)_i = \alpha + \beta \log w_i + \alpha^* D_i + \varepsilon_i \tag{B}$$

$$\log(V/L)_i = \alpha + \beta \log w_i + \alpha^* D_i + \beta^* D_i \log w_i + \varepsilon_i. \tag{C}$$

The variable D_i is a dummy variable defined by

$$D_i = \begin{cases} 0 & \text{for developed countries (observations 1–8)} \\ 1 & \text{for underdeveloped countries (observations 9–16).} \end{cases}$$

Use these estimation results to answer the following questions. State carefully all steps in your test procedures.

(a) Use the estimation results for model A over the full sample of 16 countries to test the hypothesis that the elasticity of substitution between capital and labour is equal to one. Based on your test conclusion, would it be reasonable to use a Cobb-Douglas production function to describe this industry?

(b) Answer the previous question, this time using the results for model B.

(c) Consider model B.

i. Explain briefly the role of the dummy variable term in this equation. What is its purpose? Use a diagram where appropriate.

ii. Use the estimation results for this model to test the hypothesis that developed and underdeveloped countries have the same marginal productivity condition.

iii. Use an F test to test the same hypothesis you've just examined in (b). For the purposes of this test, which model is the "restricted" model. Indicate any relationship between the value of the F statistic you obtain and the value of your test statistic in part (b).

(d) Consider model B and the hypothesis $\beta = 1$.

i. How many restrictions does this hypothesis represent?

ii. Are these linear or nonlinear?

iii. Show how model B could be rewritten so that it can be estimated subject to the restriction(s).

iv. Locate this restricted model on the output. Use these results to perform an F test of the hypothesis. How does the value of your test statistic compare with that you obtained in part (b)?

(e) Consider model C and the hypothesis $\alpha^* = \beta^* = 0$.

i. How many restrictions does this hypothesis represent?

ii. Are these linear or nonlinear?

iii. Use the estimation results for this model (and an associated restricted model) to perform an F test of this hypothesis.

(f) The output includes subsample estimations of model A. Let us denote these as follows.

$$\begin{array}{ll} \log(V/L)_i = \alpha_1 + \beta_1 \log w_i + \varepsilon_i & \text{(developed countries)} \\ \log(V/L)_i = \alpha_2 + \beta_2 \log w_i + \varepsilon_i & \text{(underdeveloped countries)} \end{array}$$

i. Indicate any relationship between the estimated coefficients of these subsample models and those for model C. Verify this numerically.

ii. Perform a Chow test for structural change. Be sure to indicate the following.

A. The parameter restrictions being tested. How many restrictions are there? Are they linear or nonlinear?

B. In this context, which model is the "restricted model"?

C. Indicate any relationship between the value of your test statistic and that obtained in part (e).

8. The attached computer output includes results for the following versions of the ACMS marginal productivity conditions for one of their industries:

$$\log(V/L)_i = \alpha + \beta \log w_i + \varepsilon_i \tag{A}$$

$$\log(V/L)_i = \alpha + \beta \log w_i + \alpha^* D_i + \varepsilon_i \tag{B}$$

$$\log(V/L)_i = \alpha + \beta \log w_i + \alpha^* D_i + \beta^* D_i \log w_i + \varepsilon_i. \tag{C}$$

The variable D_i is a dummy variable defined by

$$D_i = \begin{cases} 0 & \text{for developed countries (observations 1–7)} \\ 1 & \text{for underdeveloped countries (observations 8–13).} \end{cases}$$

Use these estimation results to answer the following questions. State carefully all steps in your test procedures, indicating clearly what parameter restriction is being tested in each case.

(a) Find the estimation results for model A, and use them to do the following.

   i. Test the hypothesis that these data can reasonably be described by a Leontief technology. (Use the p-value approach to testing.)

   ii. Test the hypothesis that these data can reasonably be described by a Cobb-Douglas technology.

(b) Find the estimation results for model C, and use them to do the following.

   i. Test the hypothesis that developed and underdeveloped countries share a common elasticity of substitution. (Use the p-value approach to testing.)

   ii. Test the hypothesis that developed and underdeveloped countries share a common CES production function.

(c) Consider the hypothesis that developed and underdeveloped countries both have Cobb-Douglas production functions (but that these may nevertheless be *different* Cobb-Douglas production functions). Call this model E.

   i. Show how this model may be estimated as a restricted version of model C.

   ii. Find the regression output for this model, and use it to perform a test of these restrictions.

(d) Consider model A constrained to satisfy $\beta = 1$.

   i. Show how model A could be rewritten to impose this restriction in estimation. Let us call this restricted model D.

   ii. Can you deduce the SSE for this restricted model from the available computer output? Show your work.

(e) Construct a flowchart-type diagram to show how the five models A–E relate to one another within ordered nests. Identify clearly which model you are viewing as the maintained hypothesis. For each model indicate (1) the restrictions relative to the maintained hypothesis, and (2) the SSE of that model.

(f) Suppose that model A had been estimated separately for the two groups of countries. What would have been the estimated elasticity of substitution for

   i. developed countries?

   ii. underdeveloped countries?

9. Consider the following Cobb-Douglas production function:

$$Q = \gamma K^{\beta} L^{\alpha}.$$

Answer the following questions.

(a) Show how this can be converted to the form of a multiple regression in which, given data series Q_i, K_i, and L_i, the parameters β and α are estimable as slope coefficients.

(b) Use the property that elasticities are logarithmic derivatives to derive convenient expressions for the elasticities of output Q with respect to (i) capital K, and (ii) labour L.

(c) What parameter restriction corresponds to the hypothesis of constant returns to scale (CRTS)?

(d) Show how your multiple regression can be rewritten to incorporate the CRTS restriction in such a way that the parameter α is eliminated; present it in a form that could be estimated as a least squares regression. Would you describe this restricted regression as a labour-intensive form or a capital-intensive form?

(e) Below is computer output for the restricted and unrestricted regressions. Use the output for the **restricted regression** to answer the following.

i. What is the coefficient estimate $\hat{\beta}$ and its estimated standard error $s_{\hat{\beta}}$? Use a hypothesis test to establish whether these data indicate clearly that capital exhibits diminishing marginal returns.

ii. Deduce an implied estimate of α and its standard error, justifying the values you present. Use these to perform a hypothesis test to establish whether the output elasticity of labour is clearly positive.

(f) Test the hypothesis of constant returns to scale.

10. Suppose that crop yield Y is determined by fertilizer application X_2 and rainfall X_3:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i.$$

Suppose that data on rainfall is unavailable so the researcher is forced to estimate

$$Y_i = b_1 + b_2 X_{2i} + e_i.$$

Answer the following questions.

(a) What are the implications of the omission of rainfall for the properties of b_2? Support your conclusion with an algebraic derivation.

(b) For the true model to be identified, rainfall must exhibit variation. Suppose that, for the period under study, all fields have the same rainfall. Does this affect your conclusion in (a)? Support your answer by explaining what happens to your algebraic result in this special case.

(c) Suppose that data on rainfall becomes available and it varies across fields, so that the true model can be estimated. Suppose that the t statistic for β_3 exceeds 1. Relative to the misspecified model,

i. Does the SSE increase, decrease, or remain the same?
ii. Does the R^2 increase, decrease, or remain the same?
iii. Does the $\bar{R}^2$ increase, decrease, or remain the same?

11. A swimwear retailer finds that its sales Y are determined almost entirely by seasonal factors. It defines the following dummy variables for spring, summer, fall, and winter:

$$D_{ki} = \begin{cases} 1 & \text{for season } k \ (k = 1, 2, 3, 4) \\ 0 & \text{otherwise.} \end{cases}$$

The retailer is considering four ways of modelling the seasonal determination of its sales. The first is a regression that includes all four dummies but no intercept:

$$(1) \quad Y_i = \alpha_1 D_{1i} + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \alpha_4 D_{4i} + \varepsilon_i.$$

The second is to use an intercept and three dummies:

$$(2) \quad Y_i = \beta_1 + \beta_2 D_{2i} + \beta_3 D_{3i} + \beta_4 D_{4i} + \varepsilon_i.$$

The third is to use an intercept and four dummies:

$$(3) \quad Y_i = \alpha + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 D_{3i} + \beta_4 D_{4i} + \varepsilon_i.$$

The fourth is to define the qualitative variable

$$X_i = \begin{cases} 0 & \text{for winter} \\ 1 & \text{for spring} \\ 2 & \text{for summer} \\ 3 & \text{for fall} \end{cases}$$

and estimate the regression

$$(4) \quad Y_i = \alpha + \beta X_i + \varepsilon_i.$$

Answer the following questions.

(a) Which of these regressions can be estimated?

(b) Explain any advantages or disadvantages of the fourth method.

(c) Can any of these regressions be tested as a restricted version of one of the others? If so, which?

(d) Which of the four regressions yields the easiest test of each of the following hypotheses?

i. Seasons 1 and 2 have the same mean sales.
ii. Seasons 1 and 3 have the same mean sales.
iii. Seasons 1 and 4 have the same mean sales.

(e) Consider the first two regressions. Would you describe the relationship between the two sets of coefficients as over-, exactly-, or underidentifying?

12. The accompanying computer output includes results for the following versions of one of the ACMS marginal productivity conditions:

$$\log(V/L)_i = \alpha + \beta \log w_i + \varepsilon_i$$
$$\log(V/L)_i = \alpha + \beta \log w_i + \alpha^* D_i + \beta^* D_i \log w_i + \varepsilon_i.$$

The first model is estimated over both the full sample of 15 countries, and over subsamples of developed and underdeveloped countries. The variable D_i is a dummy variable defined by

$$D_i = \begin{cases} 0 & \text{for developed countries (observations 1–9)} \\ 1 & \text{for underdeveloped countries (observations 10–15).} \end{cases}$$

Answer the following questions. State carefully all steps in your test procedures.

(a) Some of these regressions permit different elasticities of substitution for the two groups of countries.
   i. What is the elasticity of substitution for developed countries? For underdeveloped countries?
   ii. Test whether the production technology for developed countries can reasonably be described as Cobb-Douglas.
   iii. Test whether the production technology for underdeveloped countries can reasonably be described as Cobb-Douglas.

(b) Test whether developed and underdeveloped countries have the same elasticity of substitution.

(c) Test whether a common marginal productivity condition applies to the two country groups by using the following approaches to testing for structural change:
   i. the dummy variable approach;
   ii. the Chow test approach.

   Indicate any relationship between the two approaches.