14.661 Review Problems (due the day after the first class session)

Consumer Theory

1. Consider a consumer whose utility of goods (x_1, x_2) is given by:

$$u(x_1, x_2) = \alpha \ln(x_1 - \gamma_1) + (1 - \alpha) \ln(x_2 - \gamma_2),$$

where: $0 \leq \alpha \leq 1$, $x_1 > \gamma_1 \geq 0$, $x_2 > \gamma_2 \geq 0$.

(This is called a Stone-Geary utility function. The parameters γ_1 and γ_2 are sometimes thought of as "subsistence" levels of consumption, below which utility is not defined.)

   (a) Derive the first-order conditions for this consumer's optimal choices of x_1 and x_2 as a single equation involving the marginal rate of substitution between the two goods. What is the graphical interpretation of this condition?

   (b) Derive the uncompensated ("Marshallian") demand function for x_1 for this consumer. Use Roy's identity to confirm this.

   (c) Show that this consumer's *expenditure* on x_1 is a linear function of p_1 and y. Specifically,

$$p_j x_j = p_j \gamma_j + \alpha(y - p_1\gamma_1 - p_2\gamma_2); j = 1, 2. \tag{1}$$

   For this reason, the Stone-Geary utility function is sometimes called the "Linear Expenditure System." Interpret equation (1).

   (d) Derive the *expenditure function* for Stone-Geary preferences; as a reminder, the expenditure function gives spending required to attain a reference utility level, $\bar{u}$, given goods prices.

   (e) Derive the compensated demand function for x_1 for this consumer using Shephard's lemma. As a reminder, compensated demands describe the effect of prices on demand with utility held constant. Describe the compensated price response on a graph. Use this graph to show how income and substitution effects add up to produce the net (uncompensated) response to a price change with income rather than utility held constant.

2. More Duality

   (a) The Slutsky equation links compensated and uncompensated price responses (derivatives). Use the expenditure function to derive the Slutsky equation in one line.

   (b) Show directly that the Slutsky equation holds for Stone-Geary preferences.

3. Discounting

   (a) Suppose an annual investment of $\$C$ paid in each of 10 years yields a never-ending stream of benefits of $\$B$ starting in the 11th year. Define the internal rate of return for this project and explain how to use this rate to evaluate whether this investment is worth making.

   (b) Suppose a flow of $\$D$ for t years has a present value of $\$P$ when discounted at annual interest rate i. Find the equivalent continuously compounded rate.

'Metrics

1. Empirical researchers often work with Cobb-Douglas production functions like this:

$$Q_i = \gamma L_i^{\alpha} K_i^{\beta}; \quad \alpha > 0, \quad \beta > 0, \quad \gamma > 0,$$

where L_i is labor employed by firm i, K_i is the firm's capital, and Q_i is firm output.

(a) What must be true for this production function to exhibit constant returns to scale?

(b) Assuming you have the necessary data, describe at least two ways to construct a statistical test for constant returns to scale in production.

(c) Suppose you have data on a panel of firms, that is, T observations on N firms. Explain how to use this data to control for an unobserved but time-invariant variable that captures firm-specific managerial efficiency when estimating parameters β and α.

2. Regression basics

   (a) Characterize the relationship between the regression coefficient from a bivariate regression of wages on schooling, S_i, and the coefficient on S_i when the equation also includes test scores, A_i. This is the *omitted variables bias* (OVB) formula. Generalize the OVB formula to the case where both models include a vector of other covariates, X_i.

   (b) Explain how to calculate the coefficient on S_i in a model that incudes A_i using a two-step procedure in which the second step is a bivariate regression. This is called *regression anatomy.*

3. Instrumental variables and 2SLS

   (a) Consider an over-identified two-stage least squares (2SLS) estimator using $q > 1$ instruments to identify the effect of scalar endogenous variable S_i on dependent variable Y_i. Show that this over-identified 2SLS estimator is an IV estimator. What is the instrument?

   (b) Consider grouping data on Y_i and S_i by state, with group means denoted $\bar{Y}_j$ and $\bar{S}_j$, where j indexes states. Show that weighted least squares estimation of a regression of $\bar{Y}_j$ on $\bar{S}_j$, weighted by group size, n_j, can be computed by 2SLS in the original micro (ungrouped) data. What are the instruments?

4. Suppose that the probability a woman works is characterized by a latent-index model such that:

$$Y_i = 1\left(X_i'\beta > \varepsilon_i\right),$$

   where Y_i is woman i's employment status, X_i is a vector of her personal characteristics, and ε_i is an unobserved error term assumed to be distributed $N\left(0, \sigma^2\right)$ and independent of X_i.

   (a) Give the likelihood function for this model and use this function explain why parameters β and σ are not separately identified.

   (b) Consider a probit model of the effects of childbearing on women's employment status that can be written

$$Y_i = 1\left(X_i'\beta + \tau D_i > \varepsilon_i\right),$$

   where Y_i is i's employment status, D_i is a dummy variable indicating women with children, X_i is a vector of control variables like mother's schooling, and ε_i is an unobserved error term assumed to be distributed $N\left(0, \sigma^2\right)$ and independent of X_i.

      i. Give two possible formulas for marginal effects of D_i on employment rates and show that these depend only on the ratio $\frac{\beta}{\sigma}$.

      ii. Give formulas for at least two types of average marginal effects of D_i on employment rates.

5. Derive the formula for the standard error of an experimental average treatment effect, defined as the treatment-control difference in outcome means in a randomized trial with n observations. Show that if the outcome variable of interest is homoskedastic, the optimal (minimum sampling variance) experimental design sets the proportion treated equal to $1/2$.

6. ANOVA

   (a) Consider two random variables, Y and X. Show that the (population) variance of Y can be written as the variance of the conditional expectation function (CEF), $E[Y|X]$, plus the average of the variance of Y conditional on X. This is called the analysis of variance (ANOVA) formula. Interpret the two pieces of the formula.

   (b) Show that when the CEF is a linear function of X, the ratio of the variance of the CEF to the variance of Y is the R^2 for the population regression of Y on X.