Problem Set #2

ECONOMICS 141
FALL 2008

Due October 2

Useful Problems for Review / Discussion (not to turn in):

Problems 1.1, 1.3, 1.4, 1.5, 1.6, 3.1, 3.2, 3.6, 3.9, and 3.10 from Pindyck & Rubinfeld's text.

Problems to Turn In:

1. Consider the bivariate regression model *without* an intercept term:

$$Y_i = \beta \cdot X_i + \varepsilon_i, \qquad i = 1, ..., N,$$

where, as usually assumed, the $\{X_i\}$ are fixed constants with $N^{-1}\sum_i (X_i - \bar{X})^2 \equiv \hat{\sigma}_X^2 > 0$ and $E[\varepsilon_i] = 0$, $Var(\varepsilon_i) = \sigma^2$, and $Cov(\varepsilon_i, \varepsilon_j) = 0$ if $i \neq j$.

(a) Find the form of the least squares estimator $\hat{\beta}$ of β in this case, i.e., find an expression for the value $\hat{\beta}$ that minimizes

$$S(c) \equiv \sum_{i=1}^{N} (Y_i - c \cdot X_i)^2$$

over c.

(b) Show that this estimator is linear in the dependent variables $\{Y_i\}$ and unbiased, and give a formula for its variance.

(c) Now, assuming that $X_i \neq 0$ for all i and $\bar{X} \neq 0$, consider two other estimators of β:

$$\tilde{\beta} \equiv \frac{\bar{Y}}{\bar{X}}$$

and

$$\vec{\beta} \equiv \frac{1}{N}\sum_{i=1}^{N}\left(\frac{Y_i}{X_i}\right).$$

Show that these estimators are also linear in $\{Y_i\}$ and unbiased for β, calculate their variances, and show that those variances are at least as large as $Var(\hat{\beta})$. [HINT: You may use the fact that the (population or sample) average of $1/X^2$ is at least as large as the inverse of the average of X^2). In what case will all of these estimators have the same variance?

2. Suppose you are given the following (semi-fabricated) data on a typical automobile's fuel consumption (F) and automobile speed (S):

F_i (miles/gallon)	S_i (miles/hour)
10	10
18	20
25	30
29	40
30	50
28	60
25	70
22	80
18	90
15	100
11	110
8	120

with

$$\sum_i F_i = 239 \qquad \sum_i S_i = 780$$

$$\sum_i F_i^2 = 5471 \qquad \sum_i S_i^2 = 65000$$

$$\sum_i F_i \cdot S_i = 14350.$$

It is conjectured that there is a linear relationship between F_i and S_i:

$$F_i = \alpha + \beta \cdot S_i + U_i,$$

where U_i is a random error term assumed to have zero mean and constant variance σ^2.

(a) What are the least squares estimates of α and β?

(b) What are the residual sum of squares and R^2 for this regression?

(c) Plot F_i as a function of S_i and discuss the appropriateness of the specification used above. On the same graph, plot the fitted values $\hat{F}_i \equiv \hat{\alpha} + \hat{\beta} \cdot \dot{S}_i$ against S_i. Are the estimated residuals "nicely behaved," i.e., do they look independent and centered at zero?

(d) Re-estimate α and β using only the last eight observations (with $S_i \geq 50$) and calculate the R^2 of this regression.

(e) On the basis of this exercise what might you conclude about the relationship between speed and fuel economy? What might be a better model for this relationship than the one considered so far?

3. Suppose that, for a simple linear regression problem, you are given the six values of

$$N, \sum_{\iota=1}^{N} X_i, \sum_{\iota=1}^{N} X_i^2, \sum_{\iota=1}^{N} Y_i, \sum_{\iota=1}^{N} Y_i^2, \text{ and } \sum_{\iota=1}^{N} X_i \cdot Y_i.$$

The model is, as usual,

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

where the error terms $\{\varepsilon_i\}$ are assumed to be jointly normally distributed with $E(\varepsilon_i) = 0$, $Var(\varepsilon_i) = \sigma^2$, $Cov(\varepsilon_i, \varepsilon_j) = 0$ if $i \neq j$, and the regressors $X_1, ..., X_N$ are taken as fixed (nonrandom).

Give concise formulae using these six quantities for calculation of

(a) the sample means $\bar{Y}$ and $\bar{X}$;

(b) the sample variances $\hat{\sigma}_X^2$ and $\hat{\sigma}_Y^2$ and the sample covariance $\hat{\sigma}_{X,Y}$;

(c) the least squares estimators $\hat{\alpha}$ and $\hat{\beta}$;

(d) the R^2 of the regression;

(e) the residual sum of squared errors (SSE) and s_ε^2 for the regression;

(f) the standard errors of $\hat{\alpha}$ and $\hat{\beta}$;

(g) the F-statistic for the test of $H_0 : \beta = 0$.

Although there may be several possible formulae, try to make yours as simple as possible. You may use symbols for *simple* intermediate calculations, e.g., $r_{XY} = \hat{\sigma}_{XY}/\hat{\sigma}_X\hat{\sigma}_Y$, as long as you define these symbols explicitly in your answers.

4. *Short Answer:* Give a *brief* answer, explanation, and/or mathematical derivation to the three questions below.

A. "The Central Limit Theorem gives conditions under which "large" populations are approximately normal. That is, as the number of possible values of a random variable approaches infinity, the distribution of that random variable approaches a normal distribution." True or False? Explain.

B. In trying to model the demand for money as a function of interest rates (using a simple regression model), would you rather observe economic data during a period in which interest rates were relatively stable, or a period in which rates were volatile? Why?

C. Suppose it is observed that, for a set of data points $\{(X_i, Y_i)\}$ which are assumed to satisfy a simple linear regression model, the absolute value of the sample mean of Y_i is *greater* than the absolute value of the sample mean of X_i, but the sample variance of the dependent variable Y_i is *less* than the sample variance of the independent variable X_i – that is, $|\bar{Y}| > |\bar{X}|$ and $s_Y^2 < s_X^2$. What, if anything, does this imply about the absolute value of the least-squares slope coefficient estimate $\hat{\beta}$? How about the sign of the intercept term $\hat{\alpha}$?

[**Note:** the short answer questions in problem #4 appeared on exams for previous versions of this course, and similar questions – with some choice – will appear on the next midterm exam.]