Introduction to Econometrics (Econ 107)
Drake University, Fall 2007
William M. Boal

Signature: ______________________________

Printed name: ______________________________

MIDTERM EXAMINATION #2 VERSION A
"Two-Variable Regression"
October 9, 2007

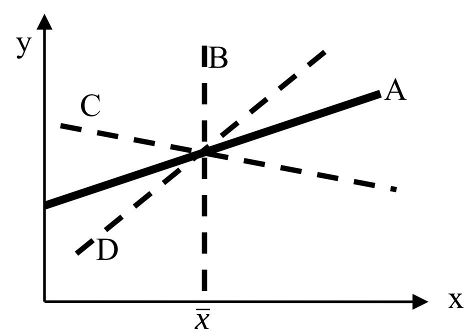
INSTRUCTIONS: This exam is closed-book, closed-notes. You may use a calculator on this exam, but not a graphing calculator or a calculator with alphabetical keys. Point values for each question are noted in brackets. A table of the t-distribution is attached.

NOTATION: In this exam, $\hat{\beta}_1$ and $\hat{\beta}_2$ denote the least-squares estimators of the intercept and slope of the line $y_i = \beta_1 + \beta_2 x_i, + \varepsilon_i$, $\hat{y}_i$ denotes a least-squares fitted value, $\hat{\varepsilon}_i$ denotes a least-squares residual, and the sample size is denoted n. The true or population value of the variance of the unobserved error term ε_i is denoted σ^2. The (unbiased) least-squares estimate of σ^2 is denoted $\hat{\sigma}^2$. The sample means of x and y are denoted $\bar{x}$ and $\bar{y}$ respectively.

I. MULTIPLE CHOICE: Circle the one best answer to each question. Feel free to use margins for scratch work [2 pts each—18 pts total]

(1) In the model $y_i = \beta_1 + \beta_2 x_i, + \varepsilon_i$, assuming $E(\varepsilon_i|x_i)=0$, the conditional mean of y (that is, $E(y_i|x_i)$) is
a. zero.
b. β_1 .
c. β_2 .
d. $\beta_2 x_i$.
e. $\beta_1 + \beta_2 x_i$.

(2) In the graph below, the solid line denoted "A" is the true population regression line. If the error term has mean zero but is *negatively correlated* with x, then the least-squares estimated line will tend to resemble
a. line A.
b. line B.
c. line C.
d. line D.
e. cannot be determined from the information given.



(3) If all data observations fitted exactly on a straight line, then the least-squares estimate of the variance of the error term would be
a. $\hat{\sigma}^2 = -1$.
b. $\hat{\sigma}^2 = 0$.
c. $\hat{\sigma}^2 = 1$.
d. $\hat{\sigma}^2 = n$.
e. None of the above.

(4) The variance of the least-squares slope estimator $\hat{\beta}_2$ is smaller, and thus the true value of β_2 is estimated more precisely,

a. the smaller the variance of the error term σ^2.
b. the smaller the variation of the x values around the sample mean $\bar{x}$.
c. the smaller the sample size.
d. All of the above.
e. None of the above.

(5) Suppose we use the least-squares predictor ($\hat{y}_{n+1} = \hat{\beta}_1 + \hat{\beta}_2 x_{n+1}$) to predict y_{n+1}. The variance of the prediction error ($y_{n+1} - \hat{y}_{n+1}$) is larger, and thus prediction is less precise,

a. the larger the variance of the error term σ^2.
b. the closer x_{n+1} is to $\bar{x}$.
c. the larger the variation of the x-values in our sample around $\bar{x}$.
d. All of the above.
e. None of the above.

(6) The variance of the prediction error tends to decrease as the sample size used for estimation increases, finally approaching

a. Var($\hat{\beta}_1$).
b. Var($\hat{\beta}_2$).
c. zero.
d. one.
e. σ^2.

(7) In time-series data, any two variables are correlated in finite sample

a. only if one variable causes the other.
b. only if neither variable causes the other.
c. if they both have trends.
d. All of the above.
e. None of the above.

(8) According to which model is the elasticity of y with respect to x equal to 0.3?

a. $y = 2.5 + 0.3\,x$.
b. $y = 2.5 + 0.3\,(1/x)$.
c. $y = 2.5 + 0.3\ln(x)$.
d. $\ln(y) = 2.5 + 0.3\,x$.
e. $\ln(y) = 2.5 + 0.3\ln(x)$.

(9) According to which model does a one-unit change in x cause approximately a four percent increase in y?

a. $y = 7.8 + 0.04\,x$.
b. $y = 7.8 + 0.04\,(1/x)$.
c. $y = 7.8 + 0.04\ln(x)$.
d. $\ln(y) = 7.8 + 0.04\,x$.
e. $\ln(y) = 7.8 + 0.04\ln(x)$.

II. MULTIPLE ANSWER: The questions below may have more than one correct answer. Write “YES” next to all correct answers and “NO” next to all incorrect answers.

(1) [5 pts] Which equations hold necessarily, regardless of the data or the model?

a. $\sum x_i \hat{\varepsilon}_i = 0$

b. $\sum (\hat{y}_i - \bar{y})^2 = \sum (y_i - \bar{y})^2 + \sum \hat{\varepsilon}_i^2$

c. $\sum x_i \hat{y}_i = 0$

d. $\sum (x_i - \bar{x})(y_i - \bar{y}) = 0$

e. $\sum \hat{\varepsilon}_i \hat{y}_i = 0$

(2) [5 pts] Which assumptions are required for the least-squares estimators to be unbiased estimators?

a. Conditional mean of error term is zero: $E(\varepsilon_i|x_i) = 0$.

b. Homoskedasticity: $Var(\varepsilon_i) = \sigma^2$, a constant.

c. No autocorrelation: $Cov(\varepsilon_i,\varepsilon_j) = 0$ for $i \neq j$.

d. Error term is normally-distributed: $\varepsilon_i \sim N(0, \sigma^2)$

e. Sample mean of x variable is zero: $\bar{x} = 0$.

(3) [5 pts] Which assumptions are required for the least-squares estimators to be method-of-moments estimators?

a. Conditional mean of error term is zero: $E(\varepsilon_i|x_i) = 0$.

b. Homoskedasticity: $Var(\varepsilon_i) = \sigma^2$, a constant.

c. No autocorrelation: $Cov(\varepsilon_i,\varepsilon_j) = 0$ for $i \neq j$.

d. Error term is normally-distributed: $\varepsilon_i \sim N(0, \sigma^2)$

e. Sample mean of x variable is zero: $\bar{x} = 0$.

(4) [5 pts] Which assumptions are required for the least-squares estimators to have the lowest variance of all unbiased estimators?

a. Conditional mean of error term is zero: $E(\varepsilon_i|x_i) = 0$.

b. Homoskedasticity: $Var(\varepsilon_i) = \sigma^2$, a constant.

c. No autocorrelation: $Cov(\varepsilon_i,\varepsilon_j) = 0$ for $i \neq j$.

d. Error term is normally-distributed: $\varepsilon_i \sim N(0, \sigma^2)$

e. Sample mean of x variable is zero: $\bar{x} = 0$.

(5) [4 pts] Suppose a demand function for milk of the form $y_i = \beta_1 + \beta_2 x_i$, is estimated by least squares. Here, y_i denotes quantity demanded in quarts and x_i denotes the price of milk in dollars. Now suppose the quantity data are converted to gallons (there are 4 quarts in a gallon).

a. $\hat{\beta}_1$ will decrease by a factor of 4.

b. $\hat{\beta}_2$ will decrease by a factor of 4.

c. The r^2 value will decrease by a factor of 4.

d. The t-statistic for $\hat{\beta}_2$ will decrease by a factor of 4.

III. PROBLEMS: Please write your answers in the boxes on this question sheet. Show your work and circle your final answers.

(1) [LS confidence intervals, tests, elasticity: 24 pts] The relationship between years of schooling and annual earnings is estimated for a sample of n=700 workers. For each worker i, let y_i denote weekly earnings and x_i denote the number of years of school completed. The model
$y_i = \beta_1 + \beta_2 x_i$ is estimated with the following results. The numbers on top are the least-squares estimates of the intercept and slope, and the numbers at the bottom in parentheses are standard errors.

Weekly earnings	=	- 1045 (45.0)	+	165 (60.0)	Years of schooling completed

a. [3 pts] Suppose a worker had 12 years of schooling. According to these results, what would be this worker's predicted weekly earnings?

b. [3 pts] Suppose a worker acquired two additional years of schooling. By how much would earnings increase? That is, what is the predicted change Δy ?

c. [3 pts] Suppose the mean years of schooling in the sample is 13 and the mean weekly earnings is $1100. (That is, $\bar{y}$ = 1100. and $\bar{x}$ = 13.) Compute the estimated elasticity of earnings with respect to years of schooling at the sample means.

d. [6 pts] Compute a **95%** confidence interval for the **intercept, β_1** .

e. [9 pts] Test the hypothesis that schooling affects earnings, against the null hypothesis that schooling has no effect (a **two-tailed test**) at **5%** significance. Give the value of the test statistic, the critical points from a table, and your conclusion (whether you can reject null hypothesis).

Value of test statistic = ________________. Critical point(s) = __________________.

Reject null hypothesis? ______________________.

(2) [LS confidence intervals, prediction: 24 pts] The relationship between total cost and units of output is measured using a sample of 15 factories. For each factory i, let y_i denote its total cost and let x_i denote its output level in. The model $y_i = \beta_1 + \beta_2 x_i$ is estimated with the following results. The numbers on top are the least-squares estimates of the intercept and slope, and the numbers at the bottom in parentheses are standard errors. Assume the error term is **normally distributed**.

Total cost	=	3600	+	55.0	Output quantity
		(30.0)		(25.0)	

a. [3 pts] Suppose a factory's output level rose from 300 units to 400 units. By how much would its total cost increase? That is, what is the predicted change Δy ?

b. [3 pts] What are the "degrees of freedom" for these estimates? Give an integer answer.

c. [6 pts] Compute a **95%** confidence interval for the **slope, β_2** .

We wish to predict cost when output is 1000 units. So we first transform the output (x) data, subtracting 1000 from every observation. Then we re-estimate the equation with the following results. The estimated variance of the error term is $\hat{\sigma}^2$ = 900.

Total cost	=	58600	+	55.0	Transformed
		(40.0)		(25.0)	output quantity

d. [3 pts] Predict cost (y) when output is 1000 units.

e. [3 pts] Compute the standard error of prediction error.

f. [6 pts] Compute a 95% prediction interval for cost when output is 1000 units.

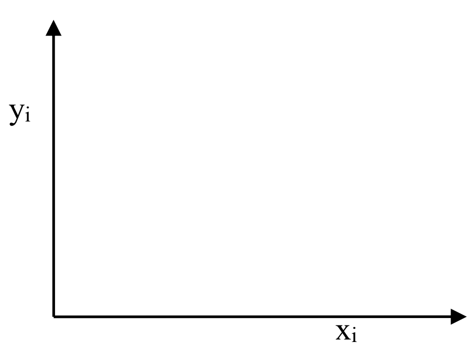
IV. CRITICAL THINKING: [10 pts] A researcher wants to estimate the effect of building codes on storm damage. Building codes are legal restrictions on the designs and permissible materials for houses, office buildings, stores, and other structures. Many people have suggested that restrictive building codes can reduce storm damage by ensuring that buildings are better-constructed.

Using a random sample of 200 counties across the United States, the researcher estimates the equation $y_i = \beta_1 + \beta_2 x_i$, where y_i denotes an index of insurance claims for storm damage in county i, and x_i denotes an index of the strictness of building codes in county i. The researcher expects to find that β_2 is negative—that is, that strict building codes reduce storm damage. However, the researcher is shocked to find that the least squares estimator of β_2 is estimated as *positive*--restrictive building codes seem to *increase* storm damage! The researcher asks your help in diagnosing the econometric problem.

a. Given these strange econometric results, which of the following assumptions do you suspect is violated in these data? Why?
   - Conditional mean of error term is zero: $E(\varepsilon_i|x_i) = 0$.
   - Homoskedasticity: $Var(\varepsilon_i) = \sigma^2$, a constant.
   - No autocorrelation: $Cov(\varepsilon_i,\varepsilon_j) = 0$ for $i \neq j$.

b. Will this violation cause the least squares estimator of β_2 to be *biased upward, biased downward, or remain unbiased?* Why? Explain your answer with a graph.

c. What should the researcher do to remedy the problem?



[end of exam]