Introduction to Econometrics (Econ 107)
Drake University, Spring 2008
William M. Boal

Signature: ______________________________

Printed name: ______________________________

MIDTERM EXAMINATION #3 VERSION B
"Multiple Regression With Cross-Section Data"
April 3, 2008

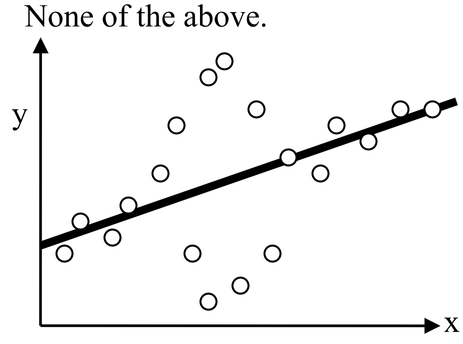
INSTRUCTIONS: This exam is closed-book, closed-notes. You may use a calculator on this exam, but not a graphing calculator or a calculator with alphabetical keys. Point values for each question are noted in brackets. Tables of the t-distribution, the F-distribution, and the chi-square distribution are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

NOTATION: In this exam, $\hat{\beta}_j$ denotes the least-squares coefficient estimators of the line $y_i = \beta_1 + \beta_2 x_{i2} + \ldots + \beta_K x_{iK} + \varepsilon_i$. The least-squares fitted value is denoted $\hat{y}_i$. The least-squares residual is denoted $\hat{\varepsilon}_i$. 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 estimator of σ^2 is denoted $\hat{\sigma}^2$. The sample mean of y is denoted $\bar{y}$.

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 graph below, the solid line is the true population regression line and the circles are observations in the sample. Which assumption appears to be violated in this sample?

a. $E(\varepsilon_i|x_i) = 0$.
b. Homoskedasticity: $Var(\varepsilon_i) = \sigma^2$, a constant.
c. No autocorrelation: $Cov(\varepsilon_i,\varepsilon_j)=0$ for $i \neq j$.
d. All of the above.
e. None of the above.



(2) Suppose we estimate the equation

$$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3},$$

and we want to test the joint hypothesis that $\beta_2 = \beta_3 = 0$. We should reject the null hypothesis at 5% significance if

a. THE F statistic is less than its 5% critical point.
b. THE F statistic is greater than its 5% critical point.
c. THE F statistic is either less than its lower 2.5% critical point or greater than its upper 2.5% critical point.
d. the t-statistics for β_2 and β_3 are both greater in absolute value than their 5% critical points.
e. either the t-statistic for β_2 or the t-statistic for β_3 is greater in absolute value than its 5% critical point.

(3) In any regression with at least two βs, Theil's adjusted R^2 (sometimes called "$\bar{R}^2$") is

a. zero.

b. negative.
c. always greater than the ordinary R^2 value.
d. always less than the ordinary R^2 value.
e. can be greater than or less than the ordinary R^2 value.

(4) If two regressors x_{i2} and x_{i3} are *closely but not perfectly* correlated, then the least-squares estimators of their coefficients
a. will have large standard errors.
b. will be zero.
c. cannot be computed.
d. will be biased.
e. will be inconsistent.

(5) Suppose we wish to test whether $\beta_3 = \beta_4 = 0$ in the equation

$$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4}$$

using an LM test. We compute the LM test statistic by first estimating the equation

$$y_i = \beta_1 + \beta_2 x_{i2} ,$$

saving the residuals $\hat{\varepsilon}_i$, and then computing $(n R_A^2)$ from the auxiliary regression
a. $\hat{\varepsilon}_i^2 = \alpha_1 + \alpha_2 x_{i2}$.
b. $\hat{\varepsilon}_i = \alpha_1 + \alpha_2 x_{i2}$.
c. $\hat{\varepsilon}_i^2 = \alpha_1 + \alpha_2 x_{i3} + \alpha_3 x_{i4}$.
d. $\hat{\varepsilon}_i = \alpha_1 + \alpha_2 x_{i3} + \alpha_3 x_{i4}$.
e. $\hat{\varepsilon}_i^2 = \alpha_1 + \alpha_2 x_{i2} + \alpha_3 x_{i3} + \alpha_4 x_{i4}$.
f. $\hat{\varepsilon}_i = \alpha_1 + \alpha_2 x_{i2} + \alpha_3 x_{i3} + \alpha_4 x_{i4}$.

(6) The equation

$$\ln(y) = 5.8 + 0.03\, x_2 + 0.05\, x_3$$

implies that, holding x_2 constant, a one-unit increase in x_3 will cause y to increase by about
a. 5.8 units.
b. 5 units.
c. 0.05 units.
d. 0.05 percent.
e. 5 percent.

(7) The equation

$$y = 5.2 + 1.3\, x_2 + 2.7\, x_3 + 0.3\, x_2^2$$

implies that, holding x_3 constant, a one-unit increase in x_2 will cause y to increase by about
a. 5.2 units.
b. 1.3 units.
c. 2.7 units.
d. $(1.3 + 0.3\, x_2)$ units.
e. $(1.3 + 0.6\, x_2)$ units.

(8) Suppose Q = quantity demanded, P = price of the good, and I = consumer income. In which specification does 0.9 equal the income elasticity of demand?
a. Q = 82.6 – 0.5 ln(P) + 0.9 ln(I) .
b. Q = 82.6 – 0.5 P + 0.9 I .
c. Q = 82.6 – 0.5 (P/I) .
d. ln(Q) = 4.3 – 0.5 P + 0.9 I .
e. ln(Q) = 4.3 – 0.5 ln(P) + 0.9 ln(I) .

(9) Suppose we wish to estimate the effect of education on earnings using data on individual workers. Moreover, we want to allow the intercept to be different for each category of worker. If we have three categories of workers, we need
a. one dummy variable.
b. two dummy variables.
c. three dummy variables.
d. four dummy variables.
e. five dummy variables.

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) [6 pts] If we estimate the equation $y_i = \beta_1 + \beta_2 x_{i2}, + \beta_3 x_{i3} + \varepsilon_i$ by ordinary least squares, then which of the following sums must necessarily equal zero, regardless of the data?

a. $\sum x_{i3}\hat{y}_i$		d. $\sum \hat{y}_i \hat{\varepsilon}_i$	
b. $\sum \hat{\varepsilon}_i$		e. $\sum y_i \hat{\varepsilon}_i$	
c. $\sum x_{i2}\hat{\varepsilon}_i$		f. $\sum \hat{\varepsilon}_i^2$	

(2) [4 pts] The variance of the least-squares slope estimator $\hat{\beta}_j$ is smaller, and thus the true value of β_j is estimated more precisely,

a. the smaller the sample size n.	
b. the larger the variance of the error term σ^2 .	
c. the larger the variation of the x_{ij} values around the sample mean $\bar{x}_j$.	
d. the less closely correlated x_{ij} is with the other regressors.	

(3) [4 pts] The least-squares estimators for the slope coefficients (the βs) will be biased and inconsistent if

a. a regressor (x) is measured with error.	
b. the error term (ε) is correlated with a regressor (x).	
c. the error term (ε_i) is heteroskedastic.	
d. the dependent variable (y) is measured with error.	

III. PROBLEMS: Write your answers in the boxes on this question sheet.

(1) [Adding regressors: 5 pts] Suppose we estimate the equation $y_i = \beta_1 + \beta_2 x_{i2}$ and then estimate $y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3}$. What are the consequences of adding the regressor x_{i3} ? Write either "must increase," or "must decrease," or "can either increase or decrease" in each box below.

a. The R^2 value	
b. Theil's adjusted R^2 (sometimes called "$\overline{R}^2$").	
c. The estimated coefficients	
d. The standard errors	
e. The sum of squared residuals	

(2) [Analysis of variance table, R^2, F-test: 20 pts] A regression program computed the following analysis-of-variance (ANOVA) table:

	Degrees of freedom ("DF")	Sums of squares ("SS")	Mean squares ("MS")
Regression (or "Model" or "Explained")	3	60	20
Residual (or "Error")	24	48	2
Total	27	108	4

a. What is the sample size?	
b. How many β coefficients were estimated, including the intercept?	
c. What is the unbiased estimate of the variance of the error term?	
d. Compute the value of R^2 (sometimes called the "coefficient of determination") to at least three significant digits.	
e. Compute the value of Theil's adjusted R^2 (sometimes called "$\overline{R}^2$") to at least three significant digits.	

f. [10 pts] Test the joint null hypothesis that all the coefficients except the intercept are zero (against the alternative hypothesis that at least one of these coefficients is not zero) at 5% significance. Give the value of the test statistic, its degrees of freedom, the critical point, and your conclusion (whether you can reject the null hypothesis).

Degrees of freedom in numerator = _____ Degrees of freedom in denominator = _____

Value of F statistic = _________________ Critical point = ____________________

Reject null hypothesis? ______________________.

(3) [Dummy variables and structural change: 20 pts] Suppose we wish to estimate the effect of inflation on growth, using a cross-section sample of **64** developed and developing countries.

$growth_i$	= average annual growth rate of country i.
$inflation_i$	= average annual inflation rate of country i.
d_i	= 1 if country i is a develop**ing** country, and
	= 0 if country i is a develop**ed** country.

The following four equations were estimated, with the sums of squared residuals (SSR) as shown.

[1]	$growth_i = 3.0 - 0.12\ inflation_i$	SSR=195
[2]	$growth_i = 2.8 - 0.13\ inflation_i + 0.4\ d_i$	SSR=180
[3]	$growth_i = 2.9 - 0.14\ inflation_i + 0.03\ (d_i \times inflation_i)$	SSR=187
[4]	$growth_i = 2.7 - 0.15\ inflation_i + 0.4\ d_i + 0.05\ (d_i \times inflation_i)$	SSR=120

First, consider equation [4].

a. According to equation [4], what is the slope for develop**ing** countries?

b. According to equation [4], what is the intercept for develop**ed** countries?

c. According to equation [4], what is the intercept for develop**ing** countries?

Second, assume that developing and developed countries have different intercepts. We wish to test the null hypothesis that all countries have the same slope, against the alternative hypothesis that the slope for develop**ed** countries is different from the slope for develop**ing** countries, at 5% significance.

d. Which equation, [1], [2], [3], or [4], is the *restricted* equation, representing the null hypothesis?

e. Which equation, [1], [2], [3], or [4], is the *unrestricted* equation, representing the alternative hypothesis?

f. [10 pts] Give the value of the test statistic, its degrees of freedom, the critical point, and your conclusion (whether you can reject the null hypothesis).

Degrees of freedom in numerator = _____ Degrees of freedom in denominator = _____

Value of F statistic = __________________ Critical point = _____________________

Reject null hypothesis? _______________________.

(4) [Heteroskedasticity: 12 pts] We have estimated the following equation by ordinary least squares, using total data for **50** states:

$$\text{spending on gasoline per capita} = \beta_1 + \beta_2 \text{ income per capita} + \varepsilon \, .$$

We believe that all the Gauss-Markov assumptions are satisfied, except that we fear that the error term (ε) might be heteroskedastic, with variance related to state population.

a. If the error term (ε) is heteroskedastic, are usual t statistics for the least squares estimators valid? (Answer *yes* or *no*.)

a. If the error term (ε) is heteroskedastic, are the least squares estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ consistent? (Answer *yes* or *no*.)

c. Given that the dependent variable is an average, is the variance of the error term (ε) more likely to be *positively* or *negatively* related to the state's population?

To test for heteroskedasticity, we save the least-squares residuals from the above equation and estimate the following auxiliary regression by least squares:

$$\hat{\varepsilon}^2 = \alpha_1 + \alpha_2 \text{ population } + \nu \ ,$$

where "population" is the state's population and υ is a new error term. The R^2 value from this auxiliary regression is 0.086.

d. Compute the value of the Breusch-Pagan test statistic.

e. Find the critical point in the appropriate table at 5% significance.

f. Can you reject the null hypothesis of no heteroskedasticity at 5% significance?

(5) [Weighted least squares: 6 pts] Suppose we wish to estimate the following production function using data on factories:

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i \, .$$

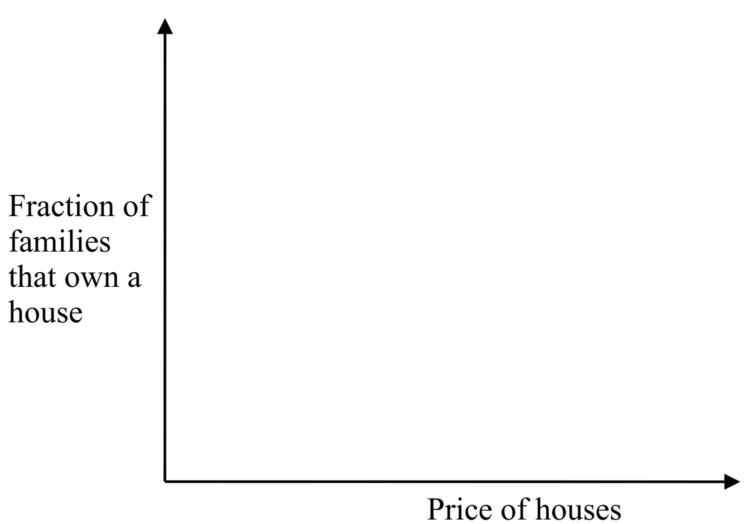
However, we believe that ε_i is heteroskedastic, with $\text{Var}(\varepsilon_i) = \alpha \text{ employ}_i$, where employ_i is the factory's employment and α is an unknown constant. The first two observations of raw data are given below at left. Transform these data to eliminate the heteroskedasticity. Put your answers in the empty boxes at right.

	Raw data				Transformed data		
i	y_i	Intercept	x_i	employ_i	y_i	Intercept	x_i
1	376	1	160	64			
2	119	1	91	49			

IV. CRITICAL THINKING: [5 pts] Suppose you want to estimate the *ceteris-paribus* effect of the price of houses on the fraction of families that own a house. Using a dataset on n=200 cities, you plan to estimate the following equation:

$$\textit{Fraction own house} = \beta_1 + \beta_2 \textit{ house price} + \varepsilon_i ,$$

where β_2 is expected to be negative from demand theory. However, suppose average family income is positively correlated with house price in your dataset. Assume that income has a positive effect on the fraction of families that own a house, but income must be omitted from the regression because income data are unavailable. Will your estimate of β_2 be biased up (toward zero), biased down (away from zero), or unbiased? Explain your reasoning using a graph. Your graph should show the true line, the pattern of observations, and the least-squares estimated line.



[end of exam]