Introduction to Econometrics (Econ 107)
Drake University, Spring 2010
William M. Boal

Signature: _____

Printed name: _____

MIDTERM EXAMINATION #3 VERSION C
"Multiple Regression With Cross-Section Data"
March 26, 2010

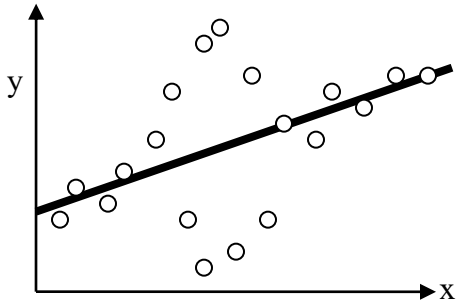
INSTRUCTIONS: This exam is closed-book, closed-notes. You may use a calculator on this exam, but not a graphing calculator or a calculator with alphabetical keys. Point values for each question are noted in brackets. Tables of the t-distribution, the F-distribution, and the chi-square distribution are 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 $\bar{y}$. The natural logarithm is denoted *ln(.)*.

I. MULTIPLE CHOICE: Circle the one best answer to each question. Use margins for scratch work [2 pts each—16 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. No autocorrelation: $Cov(\varepsilon_i,\varepsilon_j)=0$ for $i \neq j$.
c. Homoskedasticity: $Var(\varepsilon_i) = \sigma^2$, a constant.
d. All of the above.
e. None of the above.



(2) Suppose we estimate the equation

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

and we want to test the null joint hypothesis that $\beta_2 = \beta_3 = \beta_4 = 0$. We should reject the null hypothesis at 5% significance if

a. *all* the t-statistics for β_2, β_3, and β_4 are greater in absolute value than their 5% critical points.
b. THE F statistic is less than its 5% critical point.
c. THE F statistic is greater than its 5% critical point.
d. THE F statistic is either less than its lower 2.5 % critical point or greater than its upper 2.5 % critical point.
e. *any* of the t-statistics for β_2, β_3, or β_4 are greater in absolute value than their 5% critical points.

(3) Suppose we estimate the equation

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

Which is *larger*: ordinary R^2 or Theil's adjusted R^2 (sometimes called "$\bar{R}^2$") ?

a. Theil's adjusted R^2.
b. They are equal.
c. Ordinary R^2 .
d. Either the ordinary R^2 or Theil's adjusted R^2 may be larger, depending on the data.

(4) Suppose we estimate the equation

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

If x_{i2} and x_{i3} are *closely but not perfectly* correlated, then the least-squares estimators of their coefficients

a. will be biased.
b. will be inconsistent.
c. will have large standard errors.
d. will be zero.
e. cannot be computed.

(5) Suppose we wish to test the null hypothesis

$$\beta_3 = \beta_4 = 0$$

in the equation

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

using an LM test. To compute the LM test statistic, we first estimate the equation

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

saving the residuals $\hat{\varepsilon}_i$, and then estimate the auxiliary regression

a. $\hat{\varepsilon}_i^2 = \alpha_1 + \alpha_2 x_{i2}$.
b. $\hat{\varepsilon}_i^2 = \alpha_1 + \alpha_2 x_{i3} + \alpha_3 x_{i4}$.
c. $\hat{\varepsilon}_i^2 = \alpha_1 + \alpha_2 x_{i2} + \alpha_3 x_{i3} + \alpha_4 x_{i4}$.
d. $\hat{\varepsilon}_i = \alpha_1 + \alpha_2 x_{i2}$.
e. $\hat{\varepsilon}_i = \alpha_1 + \alpha_2 x_{i3} + \alpha_3 x_{i4}$.
f. $\hat{\varepsilon}_i = \alpha_1 + \alpha_2 x_{i2} + \alpha_3 x_{i3} + \alpha_4 x_{i4}$.

(6) The equation

$$\ln(y) = 3.4 + 0.15 x_2 + 0.03 x_3$$

implies that, holding x_2 constant, a one-unit increase in x_3 will cause y to increase by about

a. 0.15 units.
b. 0.03 units.
c. 0.15 percent.
d. 0.03 percent
e. 15 percent.
f. 3 percent.

(7) Suppose Q = quantity of output, L = labor input, and K = capital input. The elasticity of output with respect to labor input equals 0.81 in which equation below?

a. $Q = 3.5 + 0.81 L + 0.23 K$.
b. $Q = 3.5 + 0.81 \ln(L) + 0.23 \ln(K)$.
c. $Q = 3.5 + 0.81 (P/I)$.
d. $\ln(Q) = 3.5 + 0.81 \ln(L) + 0.23 \ln(K)$.
e. $\ln(Q) = 3.5 + 0.81 L + 0.23 K$.

(8) Suppose we wish to estimate the effect of horsepower on price using data on various models of automobile. Moreover, we want to allow the intercept to be different for each category of car (two-door, four-door, hatchback, etc.). If we have five categories of cars, we need

a. one dummy variable.
b. two dummy variables.
c. three dummy variables.
d. four dummy variables.
e. five dummy variables.

II. SHORT ANSWER: Please write your answers in the boxes on this question sheet. Use margins for scratch work.

(1) [Algebraic properties: 3 pts] Suppose we estimate the equation $y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i$ by ordinary least squares. Which equations hold necessarily, regardless of the data or the model? Write "TRUE" or "FALSE" in the boxes below.

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

b. $\sum x_{i3} y_i = 0$

c. $\sum(x_{i2} - \bar{x}_2) = 0$

(2) [Properties: 4 pts] Which of the following conditions cause the least-squares estimators for the slope coefficients (the $\hat{\beta}$s) to be biased and inconsistent? Write "YES" or "NO."

a. The dependent variable (y) is measured with error.

b. A regressor (x) is measured with error.

c. The error term (ε) is correlated with a regressor (x).

d. The error term (ε_i) is heteroskedastic.

(3) [Variance of LS estimators: 4 pts] Suppose we estimate the equation $y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i$. Assume that the Gauss-Markov assumptions hold. How would each of the following affect the *variance* of the least-squares slope estimator $\hat{\beta}_2$? Write "INCREASES VARIANCE" or "DECREASES VARIANCE" in each box below

a. The variance of the error term $\text{Var}(\varepsilon_i) = \sigma^2$ increases.

b. The variation of the x_{i2} values around their sample mean $\bar{x}_2$ increases.

c. Regressor x_{i2} is more closely correlated with regressor x_{i3}.

d. The sample size n increases.

(4) [Adding regressors: 5 pts] Suppose we first estimate the equation $y_i = \beta_1 + \beta_2 x_{i2}$ by least squares, 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}? In each box below, write one of the following:

- "*must increase*, "
- "*must decrease*, "
- "*can either increase or decrease,* "
- "*must remain constant*."

a. The standard errors of the estimated coefficients…	
b. The sum of squared residuals…	
c. The R^2 value…	
d. Theil's adjusted R^2 (also called "$\bar{R}^2$")...	
e. The estimated coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$…	

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

(1) [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")	6	488	81
Residual (or "Error")	61	122	2
Total	67	610	9.1

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? _____.

(2) [Dummy variables and structural change: 20 pts] Suppose we wish to estimate the effect of prior test scores on current test scores, using a sample of of **62** students.

$test_i$ = current test score for student i.
$old\ test_i$ = prior test score for student i.
d_i = 1 if student attends a charter school.
= 0 if student attends a regular public school.

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

[1] $test_i = 22 + 0.7\ oldtest_i$ SSR=195
[2] $test_i = 19 + 0.6\ oldtest_i + 4.0\ d_i$ SSR=180
[3] $test_i = 21 + 0.7\ oldtest_i - 0.2\ (d_i \times oldtest_i)$ SSR=187
[4] $test_i = 18 + 0.8\ oldtest_i + 6.0\ d_i - 0.3\ (d_i \times oldtest_i)$ SSR=120

First, consider equation [4].

a. According to equation [4], what is the slope for regular public-school students?
b. According to equation [4], what is the slope for charter-school students?
c. According to equation [4], what is the intercept for charter-school students?

Second, test the null hypothesis that all students have the same intercept and slope, against the alternative hypothesis that all students have the same slope but the intercepts are different for regular public-school students and charter-school students, at 5% significance. Assume the Gauss-Markov assumptions are satisfied and the error term is normally-distributed.

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? _____.

(3) [Heteroskedasticity: 12 pts] We have estimated the following equation by ordinary least squares, using total data for **60** countries:

$$spending\ on\ energy\ per\ capita = \beta_1 + \beta_2\ GDP\ per\ capita + \varepsilon .$$

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 country population.

a. If the error term (ε) is heteroskedastic, are usual standard errors for the least squares estimators valid? (Answer *yes* or *no*.)
b. If the error term (ε) is heteroskedastic, are the least squares estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ unbiased? (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 country'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:

$$\hat{\varepsilon}_i^2 = \alpha_1 + \alpha_2\ population_i + \nu_i$$

where "population" is the country's population and ν_i is a new error term. The R^2 value from this auxiliary regression is 0.065.

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?

(4) [Weighted least squares: 12 pts] Suppose we wish to estimate the following equation using data on states:

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

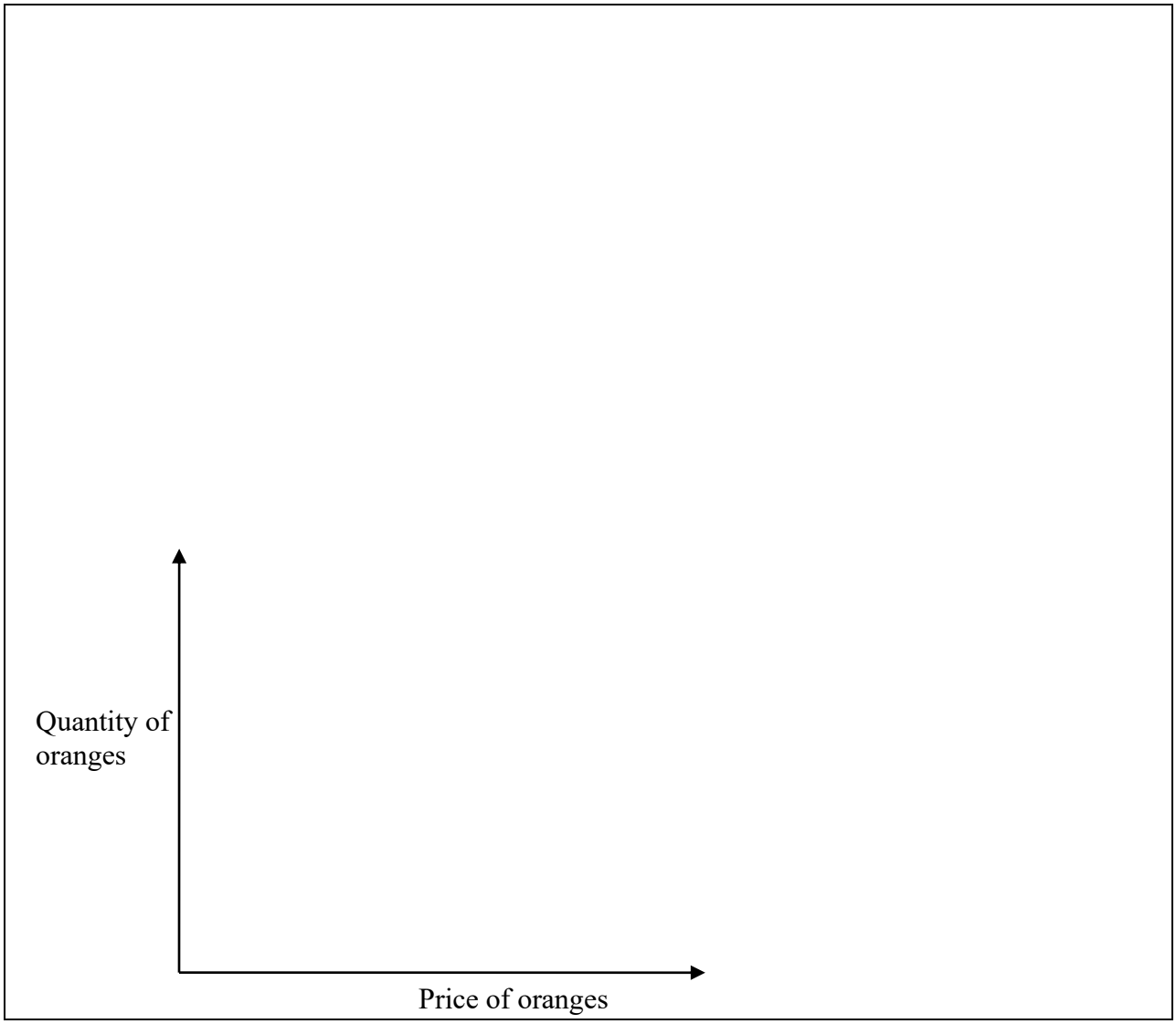
However, we believe that ε_i is heteroskedastic, with $Var(\varepsilon_i) = \alpha\ pop_i$, where pop_i is the state's population and α is an unknown constant. The first two observations of raw data are given below at left. Transform these data to eliminate the heteroskedasticity. Put your answers in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

	Raw data				Transformed data		
i	y_i	Intercept	x_i	pop_i	y_i	Intercept	x_i
1	75	1	30	9			
2	56	1	22	4			

IV. CRITICAL THINKING: [4 pts] Suppose you want to estimate the *ceteris-paribus* effect of the price of oranges on the quantity demanded of oranges. Using a dataset on n=50 cities, you plan to estimate the following demand equation:

$$\textit{Quantity of oranges} = \beta_1 + \beta_2 \textit{ price of oranges} + \varepsilon_i ,$$

where the true value of β_2 is *negative* by the law of demand. Now suppose oranges and grapefruits are substitutes in demand, and the price of oranges is positively correlated with the price of grapefruits. If the price of grapefruits is omitted from the regression equation, will the least-squares estimate of β_2 be biased *up* (closer to zero), biased *down* (too negative), or unbiased? Explain your answer. Draw a graph showing the true *ceteris*-paribus relationship between the quantity of oranges and the price of oranges as a solid line, the likely pattern of observations as dots or circles, and the least squares line as a dotted line.



[end of exam]