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

Signature: _____

Printed name: _____

# FINAL EXAMINATION VERSION B
# May 13, 2008

INSTRUCTIONS: This exam is closed-book, closed-notes. You may use a calculator for 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, F distribution, and chi-square distribution are attached.

**I. MULTIPLE CHOICE:** Circle the one best answer to each question. Feel free to use margins for scratch work [1 pt each—6 pts total]

(1) Suppose we wish to fit the equation $y = \beta_1 + \beta_2 x$ to data by the method of least squares. This method minimizes which function of the data?

a. $f(\beta_1, \beta_2) = \sum (y_i - \beta_1 - \beta_2 x_i)$.

b. $f(\beta_1, \beta_2) = \sum \left(y_i^2 - (\beta_1 + \beta_2 x_i)^2\right)$.

c. $f(\beta_1, \beta_2) = \sum (y_i - \beta_1 - \beta_2 x_i)^2$.

d. $f(\beta_1, \beta_2) = \sum |y_i - \beta_1 - \beta_2 x_i|$.

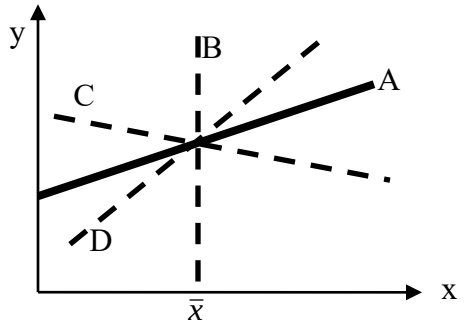
e. $f(\beta_1, \beta_2) = \sum (\beta_1 + \beta_2 x_i)^2$.

(2) Suppose the p-value for a test statistic is 0.067. If the size of the test is 5 percent, we

a. can reject the null hypothesis.
b. cannot reject the null hypothesis.
c. cannot compute the test statistic.
d. answer cannot be determined from the information given.

(3) In the graph below, the solid line denoted "A" is the true population regression line. If the error term has mean zero but is *positively 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.




(4) According to which model does a one-unit change in x cause a 0.05-unit increase in y?

a. $y = 4.5 + 0.05\ x$ .
b. $y = 4.5 + 0.05\ (1/x)$ .
c. $y = 4.5 + 0.05\ \ln(x)$ .
d. $\ln(y) = 4.5 + 0.05\ x$ .
e. $\ln(y) = 4.5 + 0.05\ \ln(x)$ .

(5) 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 four 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.

(6) If $y_t$ and $x_t$ are two independent random walks, then a regression of $y_t$ on $x_t$ will typically produce
a. an excessively large (in absolute value) t-statistic for the coefficient of $x_t$.
b. a valid t statistic for the coefficient of $x_t$ .
c. an R-square value close to zero.
d. an excessively small (in absolute value) t-statistic for the coefficient of $x_t$.

---

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

(2) [5 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 assumptions are required for the least-squares estimators to be consistent estimators?

| | |
|---|---|
| a. Conditional mean of error term is zero: $E(\varepsilon_i \mid x_{i,2}, x_{i,3}) = 0$. | |
| b. No autocorrelation: $Cov(\varepsilon_i, \varepsilon_j) = 0$ for $i \neq j$ . | |
| c. Variance of error term is zero: $Var(\varepsilon_i) = 0$. | |
| d. Error term is normally-distributed: $\varepsilon_i \sim N(0, \sigma^2)$ | |
| e. Homoskedasticity: $Var(\varepsilon_i) = \sigma^2$, a constant. | |

---

**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 average test score and spending per pupil is estimated for a sample of n=400 school districts. For each district i, let $y_i$ denote average test score and $x_i$ denote spending per pupil (in thousands of dollars). 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.

| | | | | | |
|---|---|---|---|---|---|
| Average test score | = | 58.5 (3.5) | + | 1.95 (0.6) | Spending per pupil in thousands |

a. [3 pts] Suppose a district spends \$10 thousand per pupil (that is, x = 10). According to these results, what would be this district's predicted average test score?

b. [3 pts] Suppose a district increased its spending by \$2 thousand. By how much would its average test score increase? That is, what is the predicted change $\Delta y$ when $\Delta x = 2$?

c. [3 pts] Suppose the sample mean spending per pupil in this dataset is \$10 thousand and the sample mean test score is 78. (That is, $\bar{y} = 78$. and $\bar{x} = 10$.) Compute the estimated elasticity of average test score with respect to spending per pupil at the sample means.

d. [6 pts] Compute a **95%** confidence interval for the **intercept, $\beta_1$** .

e. [9 pts] Test the hypothesis that spending has a positive effect on test scores, against the null hypothesis that spending has no effect (a **one-tailed test**) at **5%** significance. Give the value of the test statistic, the critical point(s) from a table, and your conclusion (whether you can reject null hypothesis).

Value of test statistic = _____. Critical point(s) = _____.

Reject null hypothesis? _____.

(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") | 2 | 180 | 90 |
| Residual (or "Error") | 20 | 260 | 13 |
| Total | 22 | 440 | 20 |

a. What is the sample size?

b. How many $\beta$ 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) [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 standard errors 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 $\hat{\beta}_2$ still 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 $\upsilon$ is a new error term. The $R^2$ value from this auxiliary regression is 0.092.

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) [Breusch-Godfrey test: 10 pts] Suppose we have estimated the regression model

$$y_t = \beta_1 + \beta_2 x_t + \beta_3 y_{t-1} + \varepsilon_t$$

using 91 observations, but we fear that $\varepsilon_t$ might have first-order serial correlation. Accordingly, we must estimate an auxiliary regression

a. What should be the dependent variable of the auxiliary regression?

b. What should be the regressors of the auxiliary regression?

(Note that this auxiliary regression must be estimated on observations 2 through 91 of the original data.) The $R^2$ value from this auxiliary regression is 0.053 . Test the null hypothesis of no serial correlation at 5% significance.

c. Compute the value of the Breusch-Godfrey test statistic.

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

e. Can you reject the null hypothesis at 5% significance?

(5) [Forecasting, forecast interval: 12 pts] We wish to use the equation

$$y_t = \beta_1 + \beta_2 x_{t-1} + \beta_3 y_{t-1} + \varepsilon_t$$

to forecast $y_{201}$. The last observations in our sample are $y_{200}=27$ and $x_{200}=14$. To compute the forecast, we first transform the data.

a. [2 pts] Should the dependent variable be transformed? If so, how?

b. [2 pts] Should the regressors be transformed? If so, how?

Suppose the equation has been estimated on the ***transformed data*** with the following results, standard errors in parentheses. The estimated variance of the error term is $\hat{\sigma}^2 = 19.24$.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $y_t$ | = | 37.0 | + | 0.41 | $x_{t-1}$ | + | 0.29 | $y_{t-1}$ |
| | | (2.4) | | (0.27) | | | (0.09) | |

c. [2 pts] Compute the forecast of $y_{201}$.

d. [2 pts] Compute the standard error of forecast error.

e. [4 pts] Compute a 95% forecast interval for $y_{201}$.

**IV. CRITICAL THINKING:** [5 pts] Suppose you want to estimate the *ceteris-paribus* effect of cable television on cancer. Using annual time-series data for the United States from 1958 to 2008, you estimate the following regression:

$$\textit{New cancer cases}_t = \beta_1 + \beta_2 \textit{Number of cable subscribers}_t + \varepsilon_t ,$$

Using least-squares, you find a positive estimate for $\beta_2$ , significantly different from zero at 1 percent. So the number of new cancer cases is clearly *positively correlated* with the number of cable subscribers in this dataset. Is this evidence that televisions *cause* cancer? If yes, explain why. If no, explain why not and suggest a better way to estimate the regression equation using these time-series data.

[end of exam]