Introduction to Econometrics (Econ 107)
Drake University, Spring 2006
William M. Boal

Signature: ______________________

Printed name: ______________________

FINAL EXAMINATION VERSION B
May 10, 2006

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, F, and chi-square distributions 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 for cross-sectional samples and T for time series. 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—42 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. $f(\beta_1, \beta_2) = \sum (y_i - \beta_1 - \beta_2 x)$.
b. $f(\beta_1, \beta_2) = \sum |y_i - \beta_1 - \beta_2 x|$.
c. $f(\beta_1, \beta_2) = \sum (\beta_1 + \beta_2 x)^2$.
d. $f(\beta_1, \beta_2) = \sum (y_i - \beta_1 - \beta_2 x)^2$.
e. $f(\beta_1, \beta_2) = \sum \left(y_i^2 - (\beta_1 - \beta_2 x)^2\right)$.

(2) The correlation of any random variable with itself is necessarily
a. negative.
b. zero.
c. between negative one and one.
d. exactly one.
e. infinite.

(3) Let $\hat{\theta}$ be an estimator of an unknown population parameter θ, with the property that the probability of $\hat{\theta}$ being more than any given distance from θ shrinks to zero, as the sample size increases without bound. The estimator $\hat{\theta}$ is said to be
a. unbiased
b. asymptotically unbiased.
c. consistent.
d. linear.
e. efficient.

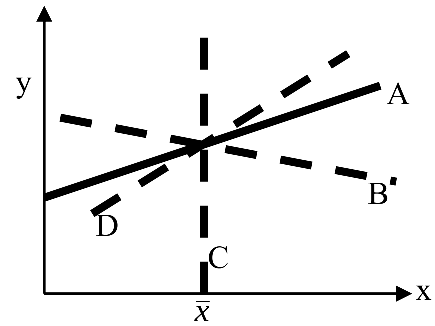
(4) If the expected value of an estimator approaches the true value of the unknown population parameter, as the sample size grows without bound, that estimator is said to be
a. unbiased.
b. asymptotically unbiased.
c. consistent.
d. linear.
e. asymptotically normal.

(5) A wider confidence interval is obtained by
a. increasing the confidence level.
b. decreasing the confidence level.
c. increasing the sample size.
d. both (b) and (c).
e. none of the above.

(6) The probability that a test will mistakenly reject the null hypothesis when it is true is called the
a. critical point of the test.
b. power of the test.
c. size or significance of the test.
d. test statistic.
e. standard error.

(7) Which equation holds necessarily, regardless of the data?
a. $\sum x_i y_i = 0$.
b. $\sum x_i \hat{y}_i = 0$.
c. $\sum y_i \hat{\varepsilon}_i = 0$.
d. $\sum (x_i - \bar{x})(y_i - \bar{y}) = 0$.
e. $\sum (y_i - \bar{y})^2 = \sum (\hat{y}_i - \bar{y})^2 + \sum \hat{\varepsilon}_i^2$.

(8) 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.



(9) The variance of the prediction error tends to decrease as the sample size used for estimation increases, finally approaching
a. σ^2.
b. zero.
c. one.
d. Var($\hat{\beta}_1$).
e. Var($\hat{\beta}_2$).

(10) In time-series data, any two variables that each have trends must be
a. causally related.
b. correlated in any finite sample.
c. orthogonal.
d. All of the above.
e. None of the above.

(11) According to which model does a one-unit increase in x cause approximately a two percent increase in y?
a. y = 5.7 + 0.02 (1/x) .
b. y = 5.7 + 0.02 ln(x) .
c. ln(y) = 5.7 + 0.02 x .
d. ln(y) = 5.7 + 0.02 ln(x) .
e. y = 5.7 + 0.02 x .

(12) Suppose a production function is estimated of the form $y_i = \beta_1 + \beta_2 x_i$, where y_i denotes output in gallons and x_i denotes labor input. Now suppose the output data are converted to liters (there are about 3.8 liters in a gallon) and the equation is re-estimated. Which of the following are true?

a. $\hat{\beta}_1$ will increase by a factor of 3.8 .
b. $\hat{\beta}_2$ will increase by a factor of 3.8 .
c. The sum of squared residuals will increase by a factor of $(3.8)^2$.
d. The r^2 value will be unaffected.
e. All of the above.

(13) Suppose we want to estimate the effect of average income on average house size. Local average house price also has an effect on house size, but we omit price from the equation for lack of data. Suppose income has a positive effect on size, price has a negative effect, and income and price are positively correlated with each other (that is, areas with high incomes also have high house prices). Then omitting price will cause the least-squares estimator of the coefficient of income

a. to be biased up (away from zero).
b. to be biased down (toward zero).
c. to be unbiased.
d. Cannot be determined from information given.

(14) 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.
b. the smaller the variance of the error term σ^2.
c. the smaller the variation of the x_{ij} values around the sample mean $\bar{x}_j$.
d. the more closely correlated x_{ij} is with the other regressors.
e. All of the above.

(15) The equation

$$\ln(wage) = 1.7 + 0.11\ educ$$

implies that if *educ* increases by one unit, then *wage* will increase by about

a. 0.11 percent.
b. 11.0 percent.
c. $0.11.
d. $1.70.
e. $11.00.

(16) Suppose Q = quantity demanded, P = price of the good, and I = consumer income. In which specification does 0.6 equal the price elasticity of demand?

a. $Q_i = 208.1 - 0.6\ P_i + 0.9\ I_i$.
b. $Q_i = 20.5 - 0.6\ (P_i/I_i)$.
c. $\ln(Q_i) = 0.7 - 0.6\ P_i + 0.1\ I_i$.
d. $Q_i = 174.3 - 0.6 \ln(P_i) + 3.1 \ln(I_i)$.
e. $\ln(Q_i) = 5.4 - 0.6 \ln(P_i) + 1.1 \ln(I_i)$.

(17) Suppose we wish to estimate the effect of income on automobile purchases using data on states:

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

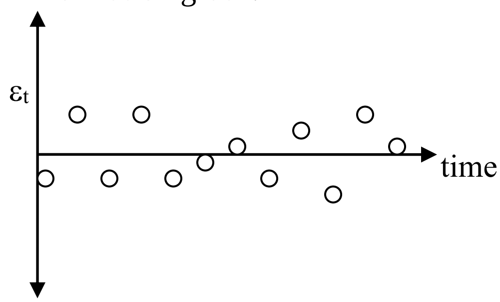
If y_i is defined as *per-capita* automobile purchases and x_i is defined as *per-capita* income in the state, then we should suspect that the variance of the error term ε_i may be

a. proportional to state population.
b. inversely proportional to state population.
c. constant and unrelated to state population.
d. zero for all observations.
e. infinite.

(18) Which is not a static model?

a. $y_t = 0.3 + 0.6\ x_t + \varepsilon_t$.
b. $y_t = 7.6 + 2.1\ x_{t,2} + 4.3\ x_{t,3} + \varepsilon_t$.
c. $y_t = 4.6 - 2.1\ x_t + \varepsilon_t$.
d. $y_t = 6.2 + 4.1\ x_{t-1} + \varepsilon_t$.

(19) The time series ε_t graphed below is
a. positively serially-correlated.
b. negatively serially-correlated.
c. serially uncorrelated.
d. Cannot be determined from the information given.



(20) By definition, if the random process u_t has a unit root,
a. $E(u_t) = 1$.
b. $Var(u_t) = 1$.
c. it does not tend to revert back to its mean.
d. $u_t = 1$.
e. the square root of $u_t = 1$.

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

II. 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: 16 pts] The relationship between electricity rates and electricity usage is estimated using a sample of n=1000 households in a variety of electric utility districts. For each household i, let y_i denote the average monthly consumption of electricity (in kilowatt-hours) and x_i denote the rate or price per kilowatt-hour. For example, if household #1 faces a rate of $0.07 per kilowatt-hour and consumes two thousand kilowatt-hours, then $x_1 = 0.07$ and $y_1 = 2000$.

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.

Electricity usage per month in kilowatt-hours	=	2880 (465)	-	10800 (3375)	Rate per kilowatt-hour

a. [2 pts] Suppose electricity were free. According to these results, how much electricity is a typical household predicted to consume per month?

kilowatt-hours

b. [2 pts] Suppose the electricity rate increased from $0.06 per kilowatt-hour to $0.10 per kilowatt-hour. By how many kilowatt-hours would electricity consumption decrease? That is, predict Δy .

kilowatt-hours

c. [2 pts] Suppose the mean electricity usage in the sample is 1800 kilowatt-hours and the mean electricity rate is $0.10 per kilowatt-hour. (That is, $\bar{y} = 1800.$ and $\bar{x} = 0.10$.) Compute the estimated elasticity of demand for electricity at the sample means.

d. [4 pts] Compute a **95%** confidence interval for the intercept.

e. [6 pts] Test the null hypothesis that electricity rates have no effect on electricity usage, against the alternative hypothesis that rates have a negative effect (a **one-tailed test**) at **5%** significance. Give the value of the test statistic, the critical point from a table, and your conclusion (whether you can reject null hypothesis).

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

Reject null hypothesis? _______________________.

(2) [Dummy variables and structural change: 22 pts] Suppose we wish to estimate the effect of tax rates on economic growth, using cross-sectional data for n=50 states. The following variables are to be used.

y_i = economic growth rate in state i.
x_i = tax rate in state i.
ds_i = 1 if state i is in the South, and 0 otherwise.
dm_i = 1 if state i is in the Midwest, and 0 otherwise.
dw_i = 1 if state i is in the West, and 0 otherwise.

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

[1]	$y_i = 0.025 – 0.007\ x_i$	SSR=360
[2]	$y_i = 0.021 + 0.002\ ds_i – 0.003\ dm_i + 0.001\ dw_i – 0.0068\ x_i$	SSR=270
[3]	$y_i = 0.019 – 0.0068\ x_i - 0.0001\ (ds_i\ x_i) + 0.0003\ (dm_i\ x_i) – 0.0005\ (dw_i\ x_i)$	SSR=280
[4]	$y_i = 0.019 – 0.0068\ x_i + 0.002\ ds_i – 0.003\ dm_i + 0.001\ dw_i$ $+ 0.0001\ (ds_i\ x_i) + 0.0002\ (dm_i\ x_i) – 0.0003\ (dw_i\ x_i)$	SSR=168

a. Although there are four official Census regions, only three dummy variables are used. If a fourth dummy variable were created for the remaining Northeast region and all four regional dummy variables were included in the same regression, then what econometric problem would arise?

b. According to equation [4], what is the intercept for the South?

c. According to equation [4], what is the slope for the Midwest?

d. According to equation [4], what is the slope for the Northeast?

Assume that all states in the sample have the same slope. We wish to test the null hypothesis that all states also have the same intercept, against the alternative hypothesis that they have different intercepts by region, at 5% significance.

e. Which equation, [1], [2], [3], or [4], is the *restricted* equation?

f. Which equation, [1], [2], [3], or [4], is the *unrestricted* equation?

g. [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) [LS tests: 6 pts] Suppose we want to test whether a year at a junior college has the same effect on a person's wage as a year at a university. We have data on 5000 individual workers. Let *yearsjc* denote years of education at a junior college, let *yearsuniv* denote years of education at a university and let *yearstot* = *yearsjc* + *yearsuniv*. We have estimated the following equation (standard errors in parentheses).

$$\underset{}{ln(wage_i)} = \underset{(0.025)}{1.75} + \underset{(0.005)}{0.012}\ yearsjc + \underset{(0.009)}{0.072}\ yearstot$$

Test the null hypothesis that a year at a junior college has the same effect on a person's wage as a year at a university, a two-tailed test, at 5% significance. Give the value of the test statistic, the critical point, and your conclusion (whether you can reject the null hypothesis).

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

Reject null hypothesis? ____.

(4) [Breusch-Godfrey test: 8 pts] Suppose we have estimated the regression

$$y_t = \beta_1 + \beta_2 x_{t,2} + \beta_3 x_{t,3} + \beta_4 y_{t-1} + \varepsilon_t$$

using 51 annual observations, but we fear that ε_t might be serially correlated. Accordingly, we have used the residuals to estimate the auxiliary regression.

$$\hat{\varepsilon}_t = \alpha_1 + \alpha_2 x_{t,2} + \alpha_3 x_{t,3} + \alpha_4 y_{t-1} + \alpha_5 \hat{\varepsilon}_{t-1} + \nu_t$$

where ν_t denotes the error term in the auxiliary regression. (Note that this auxiliary regression must be estimated on observations 2 through 51 of the original data.) The R^2 value from this auxiliary regression is 0.06 . Test the null hypothesis of no serial correlation 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 = _____

Value of test statistic = _______________ Critical point = ____________________

Reject null hypothesis? ______________________.

(5) [Forecasting: 6 pts] Suppose we have estimated the following AR(2) model.

$$y_t = 23.4 + 0.6\ y_{t-1} - 0.2\ y_{t-2} + \varepsilon_t\ .$$

where ε_t denotes an independent, identically-distributed series with $E(\varepsilon_t) = 0$ and $Var(\varepsilon_t) = \sigma^2$, constant. In our data set, $y_{T-1} = 50$ and $y_T = 40$. Compute the following forecast values.

a. Compute the forecast of y_{T+1}.

b. Compute the forecast of y_{T+2}.

c. Compute the limit of the forecast y_{T+h} as h approaches infinity.
[Hint: this is the unconditional mean of the process.]

[end of exam]