**ECN 140 - Econometrics**  
Professor Oscar Jorda  
Spring 2005

**FINAL EXAM**

Name___________________________________          Student ID:________________________

Scoring: each multiple choice questions is worth 3 points. Each correctly answered part in the analytical questions is worth 3 points. The total maximum score is 100 points although getting all answers correctly would give 102 points (i.e. you get an extra 2 bonus points!). Good luck.

Please answer your multiple choice questions in this table. Any marks done directly on the multiple choice questions themselves will not be graded.

<table>
<thead>
<tr>
<th>1.</th>
<th>2.</th>
<th>3.</th>
<th>4.</th>
<th>5.</th>
<th>6.</th>
<th>7.</th>
<th>8.</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.</td>
<td>10.</td>
<td>11.</td>
<td>12.</td>
<td>13.</td>
<td>14.</td>
<td>15.</td>
<td>----</td>
</tr>
</tbody>
</table>

**MULTIPLE CHOICE.** Choose the one alternative that best completes the statement or answers the question. Be sure to write your answers in the table provided.

1) A joint hypothesis that is linear in the coefficients and imposes a number of restrictions can be written as  
   A) $\hat{\beta} - \beta$  
   B) $(X'X)^{-1}X'Y$  
   C) $R\beta = r$  
   D) $R\beta = 0$

2) Under imperfect multicollinearity  
   A) the OLS estimator is biased even in samples of $n > 100$.  
   B) the error terms are highly, but not perfectly, correlated.  
   C) the OLS estimator cannot be computed.  
   D) two or more of the regressors are highly correlated.

3) If you reject a joint null hypothesis using the $F$-test in a multiple hypothesis setting, then  
   A) the $F$-statistic must be negative.  
   B) the regression is always significant.  
   C) all of the hypotheses are always simultaneously rejected.  
   D) a series of $t$-tests may or may not give you the same conclusion.

4) When your multiple regression function includes a single omitted variable regressor, then  
   A) the estimator for your included regressors will be biased if at least one of the included variables is correlated with the omitted variable.  
   B) the estimator for your included regressors will always be biased.  
   C) use a two-sided alternative hypothesis to check the influence of all included variables.  
   D) lower the critical value to 1.645 from 1.96 in a two-sided alternative hypothesis to test the significance of the coefficients of the included variables.

5) Sample selection bias occurs when  
   A) the availability of the data is influenced by a selection process that is related to the value of the dependent variable.  
   B) samples are chosen to be small rather than large.  
   C) data are collected from a population by simple random sampling.  
   D) the choice between two samples is made by the researcher.
6) Correlation of the regression error across observations
   A) results in correct OLS standard errors if heteroskedasticity-robust standard errors are used.
   B) results in incorrect OLS standard errors.
   C) makes the OLS estimator inconsistent, but not unbiased.
   D) is not a problem in cross-sections since the data can always be "reshuffled."

7) Misspecification of functional form of the regression function
   A) results in a type of omitted variable bias.
   B) is more serious in the case of homoskedasticity-only standard error.
   C) is overcome by adding the squares of all explanatory variables.
   D) requires alternative estimation methods such as maximum likelihood.

8) Simultaneous causality bias
   A) results in biased estimators if there is heteroskedasticity in the error term.
   B) happens in complicated systems of equations called block recursive systems.
   C) is also called sample selection bias.
   D) arises in a regression of Y on X when, in addition to the causal link of interest from X to Y, there is a causal link from Y to X.

9) The following will not cause correlation between X and u in the simple regression model:
   A) irrelevance of the regressor.       B) errors in variables.
   C) simultaneous causality.            D) omitted variables.

10) The two conditions for a valid instrument are
A) corr(Zi, Xi) = 0 and corr(Zi, ui) = 0.        B) corr(Zi, Xi) = 0 and corr(Zi, ui) ≠ 0.
C) corr(Zi, Xi) ≠ 0 and corr(Zi, ui) = 0.        D) corr(Zi, Xi) ≠ 0 and corr(Zi, ui) ≠ 0.

11) The conditions for a valid instruments do not include the following:
A) each instrument must be uncorrelated with the error term.
B) each one of the instrumental variables must be normally distributed.
C) perfect multicollinearity between the predicted endogenous variables and the exogenous variables must be ruled out.
D) at least one of the instruments must enter the population regression of X on the Z's and the W's.

12) Two stage least squares is calculated as follows: in the first stage
A) Y is regressed on the exogenous variables only. The predicted value of Y is then regressed on the instrumental variables.
B) the exogenous variables are regressed on the instruments. The predicted value of the exogenous variables is then used in the second stage, together with the instruments, to predict the dependent variable.
C) the unknown coefficients in the reduced form equation are estimated by OLS, and the predicted values are calculated. In the second stage, Y is regressed on these predicted values and the other exogenous variables.
D) the unknown coefficients in the reduced form equation are estimated by weighted least squares, and the predicted values are calculated. In the second stage, Y is regressed on these predicted values and the other exogenous variables.
13) The first difference of the logarithm of $Y_t$ equals
   A) the growth rate of $Y$ exactly.
   B) the first difference of $Y$.
   C) the difference between the lead and the lag of $Y$.
   D) approximately the growth rate of $Y$ when the growth rate is small.

14) The $j$th autocorrelation coefficient is defined as
   \[ A) \frac{\text{cov}(Y_t, X_{t-j})}{\text{var}(Y_t) \cdot \text{var}(X_{t-j})} \]
   \[ B) \frac{\text{cov}(Y_t, u_t)}{\text{var}(Y_t) \cdot \text{var}(u_t)} \]
   \[ C) \frac{\text{cov}(Y_t, Y_{t-j})}{\text{var}(Y_t) \cdot \text{var}(Y_{t-j})} \]
   \[ D) \frac{\text{cov}(Y_t, Y_{t-1})}{\text{var}(Y_t) \cdot \text{var}(Y_{t-1})} \]

15) Stationarity means that the
   A) error terms are not correlated.
   B) forecasts remain within 1.96 standard deviation outside the sample period.
   C) probability distribution of the time series variable does not change over time.
   D) time series has a unit root.

ANALYTICAL QUESTIONS. Write your answer in the space provided.

16) For this question you may assume that linear combinations of normal variates are themselves normally distributed. Let $a$, $b$, and $c$ be nonzero constants.
   
   (a) $X$ and $Y$ are independently distributed as $N(a, \sigma^2)$. What is the distribution of $(bX + cY)$?

   (b) If $X_1, \ldots, X_n$ are distributed i.i.d. as $N(a, \sigma^2_X)$, what is the distribution of $\frac{1}{n} \sum_{i=1}^n X_i$?

   (c) Draw what this distribution looks like for increasingly larger values of $n$. What is the asymptotic distribution of this statistic?
(d) Comment on the relationship between your diagram and the concept of consistency.

(e) Let $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$. What is the distribution of $\sqrt{n} (\bar{X} - a)$? Does your answer depend on $n$?

17) An extension of the Solow growth model that includes human capital in addition to physical capital, suggests that investment in human capital (education) will increase the wealth of a nation (per capita income). To test this hypothesis, you collect data for 104 countries and perform the following regression:

$$\text{RelPersInc} = 0.046 - 5.869 \times gpop + 0.738 \times sK + 0.055 \times Educ, \quad R^2 = 0.775, \quad SER = 0.1377$$

$$(0.079) \quad (2.238) \quad (0.294) \quad (0.010)$$

where $\text{RelPersInc}$ is GDP per worker relative to the United States, $gpop$ is the average population growth rate, 1980 to 1990, $sK$ is the average investment share of GDP from 1960 to 1990, and $Educ$ is the average educational attainment in years for 1985. Numbers in parentheses are for heteroskedasticity-robust standard errors.

(a) Interpret the results and indicate whether or not the coefficients are significantly different from zero. Do the coefficients have the expected sign?
(b) To test for equality of the coefficients between the OECD and other countries, you introduce a binary variable (DOECD), which takes on the value of one for the OECD countries and is zero otherwise. To conduct the test for equality of the coefficients, you estimate the following regression:

\[
\frac{\text{RelPersInc}}{\text{gpop}} = -0.068 - 0.063 \times \text{gpop} + 0.719 \times \text{sK} + 0.044 \times \text{Educ},
\]

\[
0.381 \times \text{DOECD} - 8.038 \times (DOECD \times \text{gpop}) - 0.430 \times (DOECD \times \text{sK})
\]

\[
+0.003(\text{DOECD} \times \text{Educ}), R^2=0.845, \text{SER} = 0.116
\]

Write down the two regression functions, one for the OECD countries, the other for the non-OECD countries. The \( F \)- statistic that all coefficients involving \( \text{DOECD} \) are zero, is 6.76. Find the corresponding critical value from the \( F \) table and decide whether or not the coefficients are equal across the two sets of countries.

(c) Given your answer in the previous question, you want to investigate further. You first force the same slopes across all countries, but allow the intercept to differ. That is, you reestimate the above regression but set \( \beta_{\text{DOECD} \times \text{gpop}} = \beta_{\text{DOECD} \times \text{sK}} = \beta_{\text{DOECD} \times \text{Educ}} = 0 \). The \( t \)-statistic for \( \text{DOECD} \) is 4.39. Is the coefficient, which was 0.241, statistically significant?
(d) Your final regression allows the slopes to differ in addition to the intercept. The $F$-statistic for
\[ \beta_{D_0B_0C_0D_0x_R} = \beta_{D_0B_0C_0x_R} = \beta_{D_0B_0C_0x_R} = 0 \]
is 1.05. What is your decision? Each one of the $t$-statistics is also smaller than the critical value from the standard normal table. Which test should you use?

(e) Looking at the tests in the two previous questions, what is your conclusion?
MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question. Be sure to write your answers in the table provided.

1) C
2) D
3) D
4) A
5) A
6) B
7) A
8) D
9) A
10) C
11) B
12) C
13) D
14) C
15) C

ANALYTICAL QUESTIONS. Write your answer in the space provided.

16) (a) \( E(bX + cY) = bE(X) + cE(Y) = a(b + c); \) \( \text{var}(bX + cY) = (b^2 + c^2)\sigma^2. \)

Hence \((bX+cY)\) are distributed \( N(a(b+c), \sigma^2(b^2 + c^2)) \).

(b) From (a) it follows that this is distributed as \( N(a, \frac{\sigma^2}{n}) \).

(c) The curves will be normal curves centered on \( a \), but becoming spike-like as \( n \) grows.

(d) The diagram shows that, as \( n \) grows, the probability distribution concentrates on \( a \). The probability of observing a value of \( \frac{1}{n} \sum_{i=1}^{n} X_i \) different from \( a \) becomes small as \( n \) grows. This is consistency.

(e) \( \sqrt{n}(\bar{X} - a) \) is distributed \( N(0, \sigma^2) \). This is not dependent on \( n \), in contrast to the large-sample nonnormal case where this distribution is only approached as \( n \) grows.
17) (a) A one percentage point decrease in the population growth rate increases GDP per worker relative to the United States by roughly 0.06. An increase in the investment share of 0.1 results in an increase of GDP per worker relative to the United States by approximately 0.07. For every additional year of average educational attainment, the increase is 0.055. The intercept should not be interpreted. The regression explains 77.5 percent of the variation in relative productivity. All coefficients are significantly different from zero at conventional levels. All coefficients carry the expected sign.

(b) The regression for the non-OECD countries is

\[
\text{RelPersInc} = -0.068 - 0.063\times g_{\text{pop}} + 0.719\times s_{K} + 0.044\times \text{Educ}.
\]

For the OECD countries we get

\[
\text{RelPersInc} = 0.313 - 8.101\times g_{\text{pop}} + 0.289\times s_{K} + 0.047\times \text{Educ}.
\]

The critical value is 3.32 at the 1% level and hence you can reject the null hypothesis that the coefficients are equal.

(c) Given the critical value, the coefficient is statistically significant, that is, you can reject \( \beta_{\text{OECD}} = 0 \).

(d) Given the critical value of 3.78 at the 1% level, you cannot reject the null hypothesis that the additional coefficients are all zero. The \( F \)-test is the proper procedure to use when testing for simultaneous restrictions.

(e) There is evidence that the slopes can be set equal. However, there seems to be a level difference between the two groups of countries.