Name Mr. Key
ID 7777777

ECO 5350
Intro. Econometrics

Prof. T. Fomby
Fall 2008

Mid-Term Exam

Instructions: Put your name and student ID in the upper right-hand-corner of this exam. This exam is worth a total of 105 points. The breakout of these points by question is as follows:

Q1 = 3, 2, 3, 2, 3, 3, (2, 2) = 20 points
Q2 = 2, 2, 2, 2, 2, 2, 3, 2, 2, 2, 2, 7 = 30 points
Q3 = 3, 3, 6, 4, 4 = 20 points
Q4 = (8, 2), 6, 4 = 20 points
Q5 = 3, 4, 4, 4 = 15 points

You have one hour and twenty minutes to take this test. A word from the wise: Don't get hung up on any one question. Answer the easy questions first and then go back and pick up the hard ones. Good luck.

Oh, by the way, here is a bonus question worth **3 points** for the rapt attention you have given me in my class so far. Thanks.

Question: Why does it take two Aggies to drink a bowl of soup?
+3 True or False: One has to hold his hand under the fork to catch the drippings.

statistic is "commonplace" (usually $p > 0.05$) relative to the reference distribution then we accept the null hypothesis. On the other hand, if the test statistic is "rare" (usually $p < 0.05$) we reject the null hypothesis and accept the alternative hypothesis.

1. Let's start off with some **short answer questions**.

a) What do we mean by the phrase "**in repeated samples**?" Does this phrase appear in the Classical Approach to Statistics or the Bayesian Approach? How is it applied? Give a brief explanation that a lay person in a first year class in statistics can understand.

(3) This is a classical approach term. In the classical approach to statistics we conceptualize that any random sample we take can be repeated an infinite number of times. Therefore the uncertainty of our statistics under the null hypothesis can be characterized by a reference distribution. This reference distribution is then used to determine if our observed test

b) In the first part of this course we are going to be focusing on the regression analysis of <u>cross-section</u> data. Later we will focus on the regression analysis of <u>time series</u> data. In a sequel to this course, Eco 6352, we will consider the analysis of panel data and other special regression equations involving "special" dependent variables.

(2)

c) Match up the following data types with an example of the data type.

EXAMPLES:

(3)

Real GDP observed quarterly From 1900 Q1 to 2000 QIV	A
Employment in each of the 50 States of the Union in January, 1999	C
Real Per Capita annual growth rates in 10 Countries observed from 1990 – 2000.	B

POSSIBLE DATA TYPES:

A. Time Series Data, **B.** Panel Data, **C.** Cross-Section Data

d) Define the term "Conditional Mean Function" as used in multiple regression analysis.

(2) It is the mean of the dependent variable Y given the values of the independent variables, i.e.

$$E(Y|X) = \beta_1 + \beta_2 X_2 + \cdots + \beta_K X_K.$$

e) Briefly describe state the **Gauss-Markov Theorem** and its implications for multiple regression analysis. The OLS estimators, $\hat{\beta}_1$ and $\hat{\beta}_2$, of the coefficients β_1 and β_2 ~~are~~ in the simple linear regression model are linear estimators

(3) that are unbiased, $E(\hat{\beta}_1) = \beta_1$ and $E(\hat{\beta}_2) = \beta_2$, <u>and</u> have the smallest sampling variances among the class of linear unbiased estimators. This result requires, of course, that the assumptions of the simple linear regression model. In other words the estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ are BLU.

given near exact multicollinearity we cannot determine the relative importance of the individual variables although it appears that the variables are jointly significant. This condition is detected if any of the pairwise correlation coefficients of the explanatory variables are greater than 0.9 in absolute value.

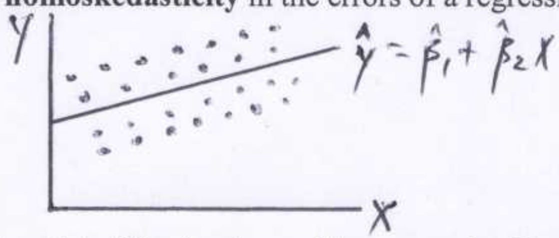
f) Briefly describe what the "**multicollinearity**" problem is in multiple linear regression and how it manifests itself and how it is detected.

(3) Exact multicollinearity exists when one of the observation vectors of the explanatory variables is a linear combination of the observation vectors of the rest of the explanatory variables. The result of this condition is that the OLS estimates cannot be computed. Near exact multicollinearity occurs when there is "almost" exact multicollinearity among the explanatory variables. A manifestation of this circumstance is that often you will find your regression equation having a significant overall F-statistic while each of the individual coefficients are statistically insignificant. This means that,

g) In the spaces below I want you to

i) Draw a scatter plot of (x,y) values with a regression line through the points that would imply **homoskedasticity** in the errors of a regression model.

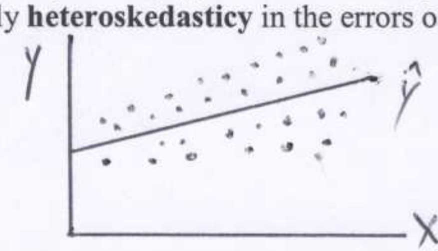
(5)



scatter is evenly distributed around regression equation

ii) Draw a scatter plot of (x,y) values with a regression line through the points that would imply **heteroskedasticy** in the errors of a regression model.

(2)



scatter is unevenly distributed around regression equation

2. Some QQ questions. (Note: Some of these questions may be slightly reworded.)

a) We discussed in class how multiple regression can be used to test the efficient market hypothesis in the field of finance. Let y denote the return on a stock over the next five years, and X1, X2, X3, and X4 denote currently available information on the financial circumstances of the firm (like debt-to-asset ratio, etc.) that supposedly underpin the value of the stock observed. The regression equation is written as

$$y = \beta_0 + \beta_1 X1 + \beta_2 X2 + \beta_3 X3 + \beta_4 X4 + \varepsilon \ .$$

Suppose that we have the above information on 100 stocks along with the firms' financial information. The null hypothesis that would test the efficient market hypothesis is

(2) H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$

(2) b) True or False. When doing tests of difference in means you use the approximate t-test when the variances of the two populations are unequal and the pooled t-test when the variances of the two populations are equal.

(2) c) By **reference distribution** for a statistical test we mean
a. the probability distribution of a test statistic in repeated sampling when assuming the null hypothesis is true
b. the probability distribution of a test statistic in repeated sampling when assuming the null hypothesis is false

(2) d) True or False. We use the calculus and the idea of second order conditions to obtain the least squares solution to the "fitting a line to a data scatter" problem.

(2) e) The statistical model $r - r_{free} = \alpha + \beta(r_{mkt} - r_{free}) + \varepsilon$ is called the
a. Standard and Poor's Model
b. Capital Asset Pricing Model
c. Efficient Market Model
d. Risk Free Interest Rate Model.

(2) f) Using the model of part e) above, a test of hypothesis of interest is
a. $H_0 : \beta = 0$ versus $H_1 : \beta \neq 0$
b. $H_0 : \beta = 0$ versus $H_1 : \beta > 0$
c. $H_0 : \alpha = 0$ versus $H_1 : \alpha \neq 0$
d. $H_0 : \alpha = 0$ versus $H_1 : \alpha > 0$

(3) g) Consider the following estimated model:

$$r - r_{free} = \underset{(0.01)}{0.06} + \underset{(0.15)}{0.95}(r_{mkt} - r_{free}) + \hat{e}$$

excess return relative to the market.

Standard errors are in the parentheses below the OLS coefficient estimates. Furthermore, assume the sample size used to get these estimates is 500. What can you say about the stock that is described by this model? Briefly explain the reasoning behind your answer.

The t-statistic on α is $t = 0.06/0.01 = 6$ whose right-tailed p-value is less than 0.05. If $t > 1.96$ we know that the right-tailed p-value is less than 0.025. The stock being analyzed provides an

(2) h) Suppose that the $\mu_x = 6$ and $\mu_y = 4$, then $E(2X + 8Y) =$ 44.

$2E(X) + 8E(Y) = 2 \cdot 6 + 8 \cdot 4$

(2) i) Assume that $\sum_1^8 X_i = 10$ and $\sum_1^8 Y_i = 108$. Then $\sum_1^8 (2X_i + 3Y_i + 4) =$ 376.

$2\sum X_i + 3\sum Y_i + 4 \cdot 8$
$= 2 \cdot 10 + 3 \cdot 108 + 32$
$= 20 + 324 + 32 = 324 + 52$
$= 376$

j) David Leonhardt thinks that there are two major causes of the current economic crisis. They are (choose two alternatives)

a. Regulators, starting with Alan Greenspan, assumed that a real estate bubble couldn't happen and that Wall Street could largely police itself.

b. The buying and selling of "naked" shorts

c. Households, struggling with incomes that haven't kept up with inflation in recent years, said yes when the lightly regulated banks offered them wishful-thinking loans.

d. Trying to finance the War in Iraq, while at the same time, running a huge government budget deficit.

2

2

k) BLU stands for Best Linear Unbiased.

l) Fill in the blanks

7

Source	SS	DF	MS	F	P-Value
Regression	18	3	6	3	0.03
Error	34	17	2		
Total	52	20			

$R^2 =$ 0.346. $R^2 = \frac{RSS}{TSS} = \frac{18}{52} = \frac{9}{26}$

26) 9.0 = .346; 78; 120; 104; 160; 156

The explanatory variables in the regression (are / are not) jointly significant. Circle the correct alternative.

Because overall F-statistic has p-value less than 0.05.

3. Consider the iso-elastic demand equation for beer and **Computer Output # 1** that you have been provided.

$$\log(Q) = \beta_1 + \beta_2 \log(PB) + \beta_3 \log(PL) + \beta_4 \log(PR) + \beta_5 \log(I) + e \quad . (1)$$

a) Are the explanatory variables in this regression equation, in total, statistically significant? Explain your answer. yes. The overall - F statistic from the ANOVA table is 29.54 which has a p-value less than 0.05. Therefore, we accept the alternative hypothesis that the explanatory variables are jointly significant

3

b) Considered individually, which of the explanatory variables are significant? Which variables are insignificant? Explain your answer.

3

$t_{lpb} = -4.27$ ($p = 0.0002$) significant lpb

$t_{lpl} = -1.04$ ($p = 0.3050$) insignificant lpl

$t_{lpr} = 2.63$ ($p = 0.0144$) significant lpr

$t_{li} = 2.22$ ($p = 50.0356$) significant li.

c) Suppose that we want to test the hypothesis that the cross-price elasticity of the remaining goods (lpr) and the income elasticity of demand for beer (li) are equal to each other. That is suppose we want to test $H_0 : \beta_4 = \beta_5$ versus $H_1 : \beta_4 \neq \beta_5$. Below construct the appropriate t-statistic for this test. You don't have to draw a conclusion, just show how you construct the statistic using the direct computation approach.

(6)

$$t = \frac{\hat{\beta}_4 - \hat{\beta}_5 - 0}{se(\hat{\beta}_4 - \hat{\beta}_5)} = \frac{0.20954 - 0.92286}{\sqrt{var(\hat{\beta}_4) + var(\hat{\beta}_5) - 2\,cov(\hat{\beta}_4, \hat{\beta}_5)}}$$

$$= \frac{0.20954 - 0.92286}{[0.00635 + 0.17265 - 2(-0.10921)]^{1/2}}$$

d) Consider the original equation in (1) above. Let $\theta = \beta_4 - \beta_5$. Show me a reparametrized form of equation (1) that would allow you to directly recover the t-statistic for the above test were OLS run on this reparametrized equation.

(4)

$$\log(Q) = \beta_1 + \beta_2 \log(PB) + \beta_3 \log(PL) + (\theta + \beta_5)\log(PR) + \beta_5 \log(I) + e$$

$$= \beta_1 + \beta_2 \log(PB) + \beta_3 \log(PL) + \theta \log(PR) + \beta_5 w + e$$

where $w = [\log(I) + \log(PR)]$. Then test $H_0: \theta = 0$ vs. $H_1: \theta \neq 0$.

e) In the SAS program you have been provided a test statement has been included that tests the equality of the cross-price elasticity of remaining goods and the income elasticity of demand for beer. What is the result of evoking the test statement? Is the cross-price elasticity of remaining goods equal to the income elasticity of demand for beer? Thoroughly explain your answer. The test statement gives us

(4)

$F = 2.53$ ($p = 0.1240$) This statistic is not significant at the 5% level and therefore we conclude that the cross-price elasticity of remaining goods is equal to the income elasticity of demand.

4. Again refer to **Computer Output # 1**. If you will notice there is a dummy variable included in the beer data set. The dummy variable is defined as follows: D = 1 when the beer is purchased during a week that contains a holiday and D = 0 otherwise.

a) Using the Additive/Multiplicative dummy variable output, fill in the following blanks:

(4)

The own price elasticity during non-holiday weeks is -0.95705.
The cross price elasticity of liquor during non-holiday weeks is -0.73015
The cross price elasticity of remaining goods during non-holiday weeks is 0.16813.
The income elasticity for beer during non-holiday weeks is 0.94566

(4) The own price elasticity during holiday weeks is $-0.95705 - 0.40368$.
The cross price elasticity of liquor during holiday weeks is $-0.73015 + 0.60520$
The cross price elasticity of remaining goods during holiday weeks is $0.16813 + 0.11306$
The income elasticity for beer during holiday weeks is $0.94566 + 0.10874$

(2) The base (reference) period of time is the (holiday / non-holiday) period. Circle the correct alternative.

b) Using the output from the Additive/Multiplicative dummy variable output, you should be able to construct a Chow test on the demand for beer between the non-holiday weeks and the holiday weeks. What is the null hypothesis of the test? What is the alternative hypothesis of the test? In the below space construct the test statistic that you would use to conduct the Chow test. (You don't have to calculate the statistic out to the final division or multiplication, just show me the form of the statistic as you would calculate it using a hand calculator.) Chow test: H_0: Demand Equations are the same

H_1: Demand Equations are different

(6)
$$F = \frac{(ESSR - ESSU)/J}{ESSU/(N-K)} = \frac{(0.08992 - 0.08428)/5}{0.08428/(30-10)}$$

c) In the Additive/Multiplicative model, I have added a test statement. What is the meaning of this test statement? What is it trying to test? What is the outcome of the test?

$H_0: \beta_6 = \beta_7 = \beta_8 = \beta_9 = \beta_{10} = 0$ [Chow test]

(4) $F = 0.27$ ($p = 0.925$) we accept the null hypothesis that the Demand Equations are the same since the p-value > 0.05

5. Now let us turn to testing for heteroskedasticity in the iso-elastic demand curve in (1) above. Consult **Computer Output #1**.

a) What can one say about OLS estimation and the OLS estimates in the presence of heteroskedasticity? when heteroskedasticity is present in the residuals of our model we can no longer use the OLS estimates and their t-values for statistical inference purposes. We must use weighted Least Squares or OLS with White's heteroskedasticity-robust standard errors.

(3)

b) What is the null hypothesis of White's heteroskedasticity test? What is the alternative hypothesis of this test?

H_0: No Heteroskedasticity in the errors of the regression model (i.e. errors are homoskedastic)

(4)

H_1: Errors are Heteroskedastic.

c) Using **Computer Output #1**, does it appear that there is heteroskedasticity in equation (1)? Why or Why not? Thoroughly explain your answer.

(4) we use the overall F-statistic in the White's heteroskedasticity "test" equation to examine the issue of hetero. Here we find that the p-value of the observed F-statistic of 0.65 is $p = 0.7311$. Therefore, we accept the null hypothesis of no heteroskedasticity.

d) Given the conclusion you drew in part c) above, was it OK to use ordinary least squares to conduct the test of equal elasticity in part c) above and the Chow test in part b) above? Explain your answer.

(4) Since it appears that we have homoskedasticity in the errors of our ~~test~~ regression equation we can conclude that conducting statistical inference by means of OLS is OK to do.

Computer Output #1

```
options nodate;

data beer;
  input q pb pl pr i d;
  datalines;
    81.7        1.78       6.95       1.11       25088    0
    56.9        2.27       7.32        .67       26561    1
    64.1        2.21       6.96        .83       25510    0
    65.4        2.15       7.18        .75       27158    0
    64.1        2.26       7.46       1.06       27162    0
    58.1        2.49       7.47        1.1       27583    1
    61.7        2.52       7.88       1.09       28235    0
    65.3        2.46       7.88       1.18       29413    0
    57.8        2.54       7.97        .88       28713    0
    63.5        2.72       7.96        1.3       30000    1
    65.9         2.6       8.09       1.17       30533    1
    48.3        2.87       8.24        .94       30373    0
    55.6           3       7.96        .91       31107    0
    47.9        3.23       8.34        1.1       31126    0
      57        3.11        8.1        1.5       32506    1
    51.6        3.11       8.43       1.17       32408    1
    54.2        3.09       8.72       1.18       33423    0
    51.7        3.34       8.87       1.37       33904    0
    55.9        3.31       8.82       1.52       34528    1
    52.1        3.42       8.59       1.15       36019    0
    52.5        3.61       8.83       1.39       34807    1
    44.3        3.55       8.86        1.6       35943    0
    57.7        3.72       8.97       1.73       37323    0
    51.6        3.72       9.13       1.35       36682    0
    53.8         3.7       8.98       1.37       38054    1
      50        3.81       9.25       1.41       36707    0
    46.3        3.86       9.33       1.62       38411    1
    46.8        3.99       9.47       1.69       38823    0
    51.7        3.89       9.49       1.71       38361    1
    49.9        4.07       9.52       1.69       41593    0
;

data beer;
  set beer;
  lq = log(q);
  lpb = log(pb);
  lpl = log(pl);
  lpr = log(pr);
  li = log(i);
  lpbd = lpb*d;
  lpld = lpl*d;
  lprd = lpr*d;
  lid = li*d;

  run;

  /* The original iso-elastic demand equation with Covariance Matrix
     of the estimates.  */

  title 'Original iso-elastic demand equation for beer' ;
```

```
proc reg data = beer;
   model lq = lpb lpl lpr li/covb;
        test lpr - li = 0;
        output out=residuals residual = res;
     run;

/* The Additive/Multiplicative Dummy Variable Equation for the iso-elastic
     demand curve for beer.  */

  title 'The Additive/Multiplicative Dummy Variable Equation' ;
proc reg data = beer;
   model lq = lpb lpl lpr li d lpbd lpld lprd lid;
          test d, lpbd, lpld, lprd, lid;
     run;

  /* The "plots" data set will allow us to produce the squared residuals for
     White's Heteroskedasticity test.  */

data plots;
    set residuals;
    res2 = res**2;

    run;

  /*  Preparing the data for White's Heteroskedasticity test.  */

data plots;
   set plots;
   lpb2 = lpb**2;
   lpl2 = lpl**2;
   lpr2 = lpr**2;
   li2 = li**2;

   run;

   /* Estimating White's Heteroskedasticity test equation */

 title 'Whites Heteroskedasticity test equation' ;
proc reg data = plots;
   model res2 = lpb lpl lpr li lpb2 lpl2 lpr2 li2;

   run;
```

The REG Procedure
Model: MODEL1
Dependent Variable: lq

Number of Observations Read	30
Number of Observations Used	30

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	0.42505	0.10626	29.54	<.0001
Error	25	0.08992	0.00360		
Corrected Total	29	0.51497			

Root MSE	0.05997	R-Square	0.8254
Dependent Mean	4.01853	Adj R-Sq	0.7975
Coeff Var	1.49242		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-3.24324	3.74300	-0.87	0.3945
lpb	1	-1.02042	0.23904	-4.27	0.0002
lpl	1	-0.58293	0.56015	-1.04	0.3080
lpr	1	0.20954	0.07969	2.63	0.0144
li	1	0.92286	0.41551	2.22	0.0356

Covariance of Estimates

Variable	Intercept	lpb	lpl	lpr	li
Intercept	14.010045406	0.6359054599	0.459996455	0.1240305592	-1.513136536
lpb	0.6359054599	0.0571410059	-0.058720912	0.0043690644	-0.055404986
lpl	0.459996455	-0.058720912	0.3137680884	-0.007871409	-0.101994071
lpr	0.1240305592	0.0043690644	-0.007871409	0.0063509159	-0.010921596
li	-1.513136536	-0.055404986	-0.101994071	-0.010921596	0.1726520356

The REG Procedure
Model: MODEL1
Dependent Variable: lq

Number of Observations Read	30
Number of Observations Used	30

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	0.43069	0.04785	11.36	<.0001
Error	20	0.08428	0.00421		
Corrected Total	29	0.51497			

Root MSE	0.06492	R-Square	0.8363
Dependent Mean	4.01853	Adj R-Sq	0.7627
Coeff Var	1.61542		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-3.23450	4.87625	-0.66	0.5147
lpb	1	-0.95705	0.32359	-2.96	0.0078
lpl	1	-0.73015	0.73987	-0.99	0.3355
lpr	1	0.16813	0.11908	1.41	0.1734
li	1	0.94566	0.51658	1.83	0.0821
d	1	-1.97744	10.57875	-0.19	0.8536
lpbd	1	-0.40368	0.66949	-0.60	0.5533
lpld	1	0.60520	1.43241	0.42	0.6772
lprd	1	0.11306	0.18720	0.60	0.5527
lid	1	0.10874	1.20978	0.09	0.9293

The REG Procedure
Model: MODEL1

Test 1 Results for Dependent Variable lq

Source	DF	Mean Square	F Value	Pr > F
Numerator	5	0.00113	0.27	0.9255
Denominator	20	0.00421		

The REG Procedure
Model: MODEL1
Dependent Variable: res2

Number of Observations Read	30
Number of Observations Used	30

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	0.00020672	0.00002584	0.65	0.7311
Error	21	0.00084000	0.00004000		
Corrected Total	29	0.00105			

Root MSE	0.00632	R-Square	0.1975
Dependent Mean	0.00300	Adj R-Sq	-0.1082
Coeff Var	211.00656		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-4.78446	17.64503	-0.27	0.7889
lpb	1	-0.06882	0.12105	-0.57	0.5757
lpl	1	1.41763	1.72003	0.82	0.4191
lpr	1	0.00231	0.00928	0.25	0.8056
li	1	0.66205	3.58627	0.18	0.8553
lpb2	1	0.04128	0.06228	0.66	0.5147
lpl2	1	-0.34432	0.40707	-0.85	0.4072
lpr2	1	0.03031	0.02086	1.45	0.1611
li2	1	-0.03266	0.17195	-0.19	0.8512