Name Mr. Key
ID

ECO 5350
Intro. Econometrics

Prof. T. Fomby
Fall 2006

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 53 points. The breakout of these points by questions is as follows:

Q1 = (a) 1 (b) 1 (c) 1 (d) 1 (e) 1 (f) 2 (g) 1 (h) 2 (i) 1 (j) 1 (k) 5 (l) 3 (m) 2 (n) 1 (o) 3
= 26 points
Q2 = (a) 3 (b) 3 = 6 points
Q3 = (a) 5 (b) 3 (c) 3 = 11 points
Q4 = 10 points

You have one hour and twenty minutes to take this test. Good luck

1. Let's start off with some QQ questions

a) The two types of probability distributions are:

1 The discrete pdf and the continuous pdf.

b) In the article I handed out in class, "Reading the Future," the author Carol Cropper reports on the research of Professors Trevino and Robertson that shows how price-earnings ratios for the Standard & Poor's stock index are related to subsequent 5-year returns on the S&P index. Professors Trevino and Robertson came up with a formula Y = 20.6 – 0.57X where Y = the subsequent 5 year annualized return on the S&P index and X = the current price-earnings ratio. Given this formula, what p-e ratio is so "high" that the subsequent 5 year annualized return on the S&P index is expected to be zero?

1 36.14. $20.6 - 0.57X = 0 \Rightarrow X = 20.6/0.57 = 36.14$

c) True or False

1 The **Gauss-Markov Theorem** states that, among all linear unbiased estimators of the coefficients of linear regression models that satisfy the assumptions SR.1 – SR.5, the Ordinary Least Squares estimators of β_1 and β_2, namely b_1 and b_2 respectively, have the smallest sampling variance.

d) True or False

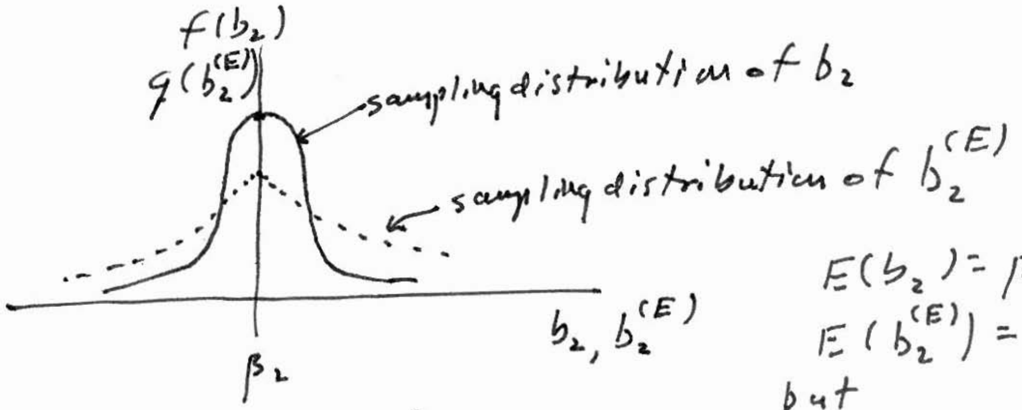
1 For a fixed number of observations on the explanatory variable x, the closer together the observations on x are, the smaller the sampling variance of b_2.

e) True or False

1 The **Extreme Values Estimator** of the slope β_2 of the population regression function $E(y\,|\,x)$ is a linear unbiased estimator but, unfortunately, it has smaller sampling variance than the Ordinary Least Squares estimator and thus is an inefficient estimator of β_2.

f) In the below space draw two sampling distributions one for the OLS estimator b_2 and one for Extreme Values Estimator, say, $b_2^{(E)}$ so as to illustrate the Gauss-Markov Theorem.

2



$E(b_2) = \beta_2$

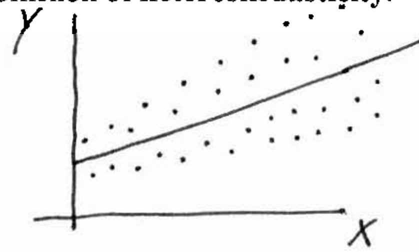
$E(b_2^{(E)}) = \beta_2$

but

$Var(b_2) < Var(b_2^{(E)})$

1 g) When the squared residuals of a simple linear regression are **not** flat as a function of x, we have (homoskedasticity /heteroskedasticity) in the errors of our model. Circle the correct alternative.

2 h) In the below space draw a plot of (x,y) points that would indicate that the data is subject to the phenomenon of **heteroskedasticity**.



$\hat{y} = b_1 + b_2 X$ one example: Increase in variability of y around the sample regression function as X becomes larger

1 i) True or False. **Weighted Least Squares** (WLS) is different from **Ordinary Least Squares** (OLS) in that WLS obtains estimates of the intercept and slope of the conditional mean function by placing less weight on the y values that have small variance and more weight on y values that have large variance whereas OLS values all of the y observations equally.

1 j) True or False. According to **Aitken's Theorem**, when heteroskedasticity is present, the WLS estimates of the intercept and slope of the conditional mean function have smaller sampling variances than the corresponding OLS estimates.

k) Consider the following **ANOVA table**. Fill in the blanks

5

Source	SS	DF	MS	F	P-Value
Regression	12	3	4	4	0.03
Error	40	40	1		
Total	52	43			

$(N-1) = 43$

3 l) The number of observations used to generate the above ANOVA table is 44. $\therefore N = 44$
The number of explanatory variables (apart from the intercept) in the above regression model is 3. The explanatory variables in the regression (are / are not) jointly significant. Circle the correct alternative.

2 m) Fill in the blanks: There are two types of prediction confidence intervals in linear regression: (1) Prediction confidence intervals for the conditional mean and (2) Prediction confidence intervals for the yet to be realized individual observation.

1 n) True or False. With respect to the confidence intervals of types (1) and (2) in the previous question, the confidence intervals of type (1) are always narrower than the confidence intervals of type (2).

o) Consider the following two models:

$$y_i = \beta_1 + \beta_2(1/x_{i2}) + e_i \quad (1)$$

3

$$\ln(y_i) = \beta_1 + \beta_2 \ln(x_i) + e_i \quad (2)$$

Model ___2___ is the model that one might use to estimate the elasticity of demand while Model ___1___ is the model that one might use to estimate the Phillips curve relationship in macroeconomics. These models are (nested / non-nested). Circle the correct choice. True or False The better of models (1) and (2) is the one that has the highest adjusted R^2.

2. **Specification Analysis.**

a) Suppose that you have estimated the following equation:

$$\widehat{educ} = 10.36 + 0.131\text{meduc}$$

where educ represents the number of years of schooling obtained by a child at the end of his/her academic career and meduc represents the number of years of education of the mother. Suppose that in addition to meduc, a second variable, feduc, the father's education in terms of the number of years of schooling should have been included in the regression but was not because no data was collected on the father's education. Furthermore, assume that the years of education of the father is positively correlated with the education of the mother. In this circumstance how would you interpret the 0.131 coefficient for the effect of the mother's education on the subsequent attainment of education by the child? Thoroughly explain your reasoning. We would expect the

3 father's education to have a positive effect on the educational attainment of the child ($\beta_2 > 0$). With $\mathrm{Cov}(X_1, X_2) > 0$ we know that the OLS estimator b_1 has a positive bias, i.e. $E(b_1) > \beta_1$. Thus, on average, a estimate like 0.131 would overstate the effect of the mother's education on the child's educational attainment

b) Consider the following housing price equation:

$$\log(\widehat{price}) = 11.08 + 0.255\text{rooms} + 0.002\text{sqrft}$$

where price is the price of a house, rooms is the number of rooms in the house and sqrft is the number of square feet in the house. Suppose that in fact sqrft is an irrelevant variable in the population regression function for housing prices and that rooms is the only relevant variable in the population regression function. Describe the properties of the above coefficient estimate of the rooms effect on price? The estimator b_1 is still unbiased

3 ($E(b_1) = \beta_1$) but is inefficient in that its sampling variance is larger than would be the case were the redundant variable sqrft dropped from the regression.

3. Consider the **Computer Output** you have been given. Recall the Alesina and Summers model where we have a regression equation that explains a developing country's future rate of inflation as a function of the independence of its central bank.

(a) Explain to me the "Alesina and Summers" hypothesis as explained in class. Does the data analysed in the SAS output support their hypothesis. Why or why not?

5

Alesina and Summers hypothesis:
The more independent the central bank is of the goverment, the lower, on average, the rate of inflation in the country. The coefficient on the central bank independence variable should be negative and, in fact, is statistically so since its p-value is < 0.0001.

(b) Suppose we have a country whose central bank independence measure is 0.75. Use the computer output to produce a prediction of the country's subsequent inflation rate and also provide me with a 95% confidence interval for your prediction. Show your work below so that you will obtain full credit for your answer.

3

$\hat{Y}_0 = 9.44019 - 1.63558(0.75) = 8.21351$

$se(\hat{Y}_0) = \sqrt{RMSE^2 + se(\hat{b}_1)^2} = \sqrt{0.89865^2 + 0.49976^2}$

1.03

$\hat{Y}_0 \pm t_{14,.025}\, se(\hat{Y}_0) \Rightarrow 8.21351 \pm 2.145(1.03) \Rightarrow (6.00, 10.42)$

(c) I have said in class that prediction confidence intervals are "random intervals." Explain to me what I meant by this comment and how it relates to 95% probability.

3

repeated samples, 95% of so constructed prediction confidence intervals will encompass the future value, Y_0.

4. Using the following data, fill in the appropriate blanks:

10

X	Y	X^2	XY
1	2	1	2
2	3	4	6
3	4	9	12
4	5	16	20

$\bar{X} = \frac{1+2+3+4}{4} = 2.5$ $\bar{Y} = \frac{2+3+4+5}{4} = 3.5$ $\bar{X}\cdot\bar{Y} = (2.5)(3.5) = 8.75$

$\Sigma Xy = 2+6+12+20 = 40$ $\Sigma X^2 = 1+4+9+16 = 30$

$$b_2 = \frac{\sum X_i Y_i - N\overline{X}\overline{Y}}{\sum X_i^2 - N\overline{X}^2} = \frac{40 - 4(8.75)}{30 - 4(2.5)^2} = \frac{40-35}{30-25} = \frac{5}{5} = 1$$

$b_1 = \overline{Y} - b_2\overline{X} = 3.5 - 1(2.5) = 1.0$

When $X_0 = 2.5$ then $\hat{Y}_0 = b_1 + b_2 X_0 = 1 + 1(2.5) = 3.5$

Suppose that the probability value for the statistic for b_2 is $\Pr(t > |t_0|) = 0.03$. Further suppose that you would like to test $H_0 : \beta_2 = 0$ versus $H_1 : \beta_2 \neq 0$ at the 5% level of statistical significance. I conclude that the explanatory variable X is statistically significant.

Computer Output

```
/* x = average index of central bank independence (1 = little independence
         4 = very independent)
   y = average Inflation 1955 - 1988
   Source: Alesina and Summers (1993) Jo. of Money,
   Credit, and Banking  */

data in;
  input x y;
cards;
1.5 8.5
1 7.6
2 6.4
1.75 7.3
2 6.7
2 6.1
2.5 6.5
2 4.1
2 6.1
2 6.1
2.5 4.5
2.5 4.2
2.5 4.9
3.5 4.1
4 3
4 3.2


proc means data=in;
   var x y;

run;

proc reg data = in;
   model y = x / r clm cli;

run;

/*  Here we use the transformed model to obtain the point
    prediction when x = 0.75 (it is the estimate of the intercept
    in the transformed model) and the ingredients for the construction
    of the standard error of the prediction error.  We use the
    Standard Error of the Transformed Regression (RMSE in SAS) or the Mean
    Square of the Error in the ANOVA table) to construct the standard error of the
    prediction error.  se(prediction error) = sqrt(RMSE^2 + se(intercept)^2)=
    sqrt(mean square error + se(intercept)^2).  */

data in;
  set in;
  xstar = x - 0.75;

proc reg data=in;
  model y = xstar;

run;
```

The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
x	16	2.3593750	0.8365044	1.0000000	4.0000000
y	16	5.5812500	1.6203780	3.0000000	8.5000000

The REG Procedure
Model: MODEL1
Dependent Variable: y

Number of Observations Read	16
Number of Observations Used	16

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	28.07829	28.07829	34.77	<.0001
Error	14	11.30609	0.80758		
Corrected Total	15	39.38437			

Root MSE	0.89865	R-Square	0.7129
Dependent Mean	5.58125	Adj R-Sq	0.6924
Coeff Var	16.10129		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	9.44019	0.69194	13.64	<.0001
x	1	-1.63558	0.27738	-5.90	<.0001

The REG Procedure
Model: MODEL1
Dependent Variable: y

Output Statistics

Obs	Dependent Variable	Predicted Value	Std Error Mean Predict	95% CL Mean		95% CL Predict		Residual
1	8.5000	6.9868	0.3276	6.2843	7.6894	4.9354	9.0383	1.5132
2	7.6000	7.8046	0.4389	6.8632	8.7460	5.6596	9.9496	-0.2046
3	6.4000	6.1690	0.2458	5.6419	6.6962	4.1708	8.1672	0.2310
4	7.3000	6.5779	0.2811	5.9749	7.1809	4.5584	8.5975	0.7221
5	6.7000	6.1690	0.2458	5.6419	6.6962	4.1708	8.1672	0.5310
6	6.1000	6.1690	0.2458	5.6419	6.6962	4.1708	8.1672	-0.0690
7	6.5000	5.3512	0.2280	4.8622	5.8403	3.3627	7.3397	1.1488
8	4.1000	6.1690	0.2458	5.6419	6.6962	4.1708	8.1672	-2.0690
9	6.1000	6.1690	0.2458	5.6419	6.6962	4.1708	8.1672	-0.0690
10	6.1000	6.1690	0.2458	5.6419	6.6962	4.1708	8.1672	-0.0690
11	4.5000	5.3512	0.2280	4.8622	5.8403	3.3627	7.3397	-0.8512
12	4.2000	5.3512	0.2280	4.8622	5.8403	3.3627	7.3397	-1.1512
13	4.9000	5.3512	0.2280	4.8622	5.8403	3.3627	7.3397	-0.4512
14	4.1000	3.7157	0.3880	2.8834	4.5479	1.6162	5.8151	0.3843
15	3.0000	2.8979	0.5075	1.8094	3.9864	0.6843	5.1114	0.1021
16	3.2000	2.8979	0.5075	1.8094	3.9864	0.6843	5.1114	0.3021

Output Statistics

Obs	Std Error Residual	Student Residual	-2-1 0 1 2	Cook's D
1	0.837	1.808	| |*** |	0.250
2	0.784	-0.261	| | |	0.011
3	0.864	0.267	| | |	0.003
4	0.854	0.846	| |* |	0.039
5	0.864	0.614	| |* |	0.015
6	0.864	-0.0799	| | |	0.000
7	0.869	1.322	| |** |	0.060
8	0.864	-2.394	| ****| |	0.232
9	0.864	-0.0799	| | |	0.000
10	0.864	-0.0799	| | |	0.000
11	0.869	-0.979	| *| |	0.033
12	0.869	-1.324	| **| |	0.060
13	0.869	-0.519	| *| |	0.009
14	0.811	0.474	| | |	0.026
15	0.742	0.138	| | |	0.004
16	0.742	0.407	| | |	0.039

Sum of Residuals	0
Sum of Squared Residuals	11.30609
Predicted Residual SS (PRESS)	13.69189

The REG Procedure
Model: MODEL1
Dependent Variable: y

Number of Observations Read	16
Number of Observations Used	16

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	28.07829	28.07829	34.77	<.0001
Error	14	11.30609	0.80758		
Corrected Total	15	39.38437			

Root MSE	0.89865	R-Square	0.7129
Dependent Mean	5.58125	Adj R-Sq	0.6924
Coeff Var	16.10129		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	8.21351	0.49976	16.44	<.0001
xstar	1	-1.63558	0.27738	-5.90	<.0001

Table 2 **Right-Tail Critical Values for the *t*-distribution**

DF	$\alpha = .10$	$\alpha = .05$	$\alpha = .025$	$\alpha = .01$	$\alpha = .005$
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
31	1.309	1.696	2.040	2.453	2.744
32	1.309	1.694	2.037	2.449	2.738
33	1.308	1.692	2.035	2.445	2.733
34	1.307	1.691	2.032	2.441	2.728
35	1.306	1.690	2.030	2.438	2.724
36	1.306	1.688	2.028	2.434	2.719
37	1.305	1.687	2.026	2.431	2.715
38	1.304	1.686	2.024	2.429	2.712
39	1.304	1.685	2.023	2.426	2.708
40	1.303	1.684	2.021	2.423	2.704
50	1.299	1.676	2.009	2.403	2.678
60	1.296	1.671	2.000	2.390	2.660
70	1.294	1.667	1.994	2.381	2.648
80	1.292	1.664	1.990	2.374	2.639
90	1.291	1.662	1.987	2.368	2.632
100	1.290	1.660	1.984	2.364	2.626
110	1.289	1.659	1.982	2.361	2.621
120	1.289	1.658	1.980	2.358	2.617
∞	1.282	1.645	1.960	2.326	2.576

Source: This table was generated using the SAS® function TINV.