Name Mr. Key
ID

ECO 5350
Intro. Econometrics

Prof. T. Fomby
Spring 2006

Mid-Term 2

Instructions: Put your name and student ID in the upper right-hand-corner of this exam. This exam is worth a total of ~~50~~ 58 points. The breakout of these points by question is as follows:

Q1 = (2,2,2) = 6 points
Q2 = (3,3) = 6 points
Q3 = (3,3,3) = 9 points — Q4 = (3,3,2) = 8 points
Q5 ~~Q4~~ = (3,3,3,3,3,3) = 18 points
Q6 ~~Q5~~ = (2,2,2,2,2) = 10 points

plus **one point** for correctly signing your name **legibly** above.

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

1. **Define the following terms:**

(2) a) **Spurious Regression** A regression involving time series variables that leads to incorrect statistical inference.

(2) b) **Multicollinearity** The case where the explanatory variables of a regression are nearly linear combinations of each other. The result is that, though the OLS estimates are unbiased their sampling variances are very large. Thus, inference, although consistent (valid), is very imprecise and not very informative.

c) **Dummy Variable**

(2) A zero-one classification variable indicating whether a given condition is present or not.

2. Consider the following SAS code.

```
proc reg data=ceo;
  model salary = ceoten ceotensq profits;

proc reg data=ceo;
  model ceoten = ceotensq profits;
     output out = result1 r=residu;
  model salary = ceotensq profits;
     output out = result2 r = residv;

data together;
   merge result1 result2;

proc reg data = together;
   model residv = residu/noint;

   run;
```

residu : The part of ceoten net of ceotensq and profits

residv : The part of salary net of ceotensq and profits

The last proc reg statement produces the following regression result:

$$residv = 49.747 residu + \hat{e}$$

(a) Using this result, the above SAS code and the **Frisch-Waugh theorem**, you should be able to fill in one of the following blanks: Fill in the appropriate blank.

(3)

Salary = _____ + 49.747 *ceoten + _____*ceotensq + _____*profits

(3) (b) Briefly describe to me the interpretation you would give for the above variable "residv." : The part of salary net of ceotensq and profits

3. We are going to conduct a **subset F-test** on the following regression equation

$$Salary_i = \beta_1 + \beta_2 ceoten_i + \beta_3 ceotensq_i + \beta_4 profits_i + e_i \quad .$$

The hypothesis we want to test is $H_0: \beta_2 = \beta_3 = 0$.

(a) In the below space, write out the restricted regression model that the above hypothesis implies.

(3) $salary_i = \beta_1 + \beta_4 profits + e_i$

(b) Suppose N = 44, $SSE_R = 65$, and $SSE_U = 50$. Define what N, SSE_R, and SSE_U represent. N = number of observations

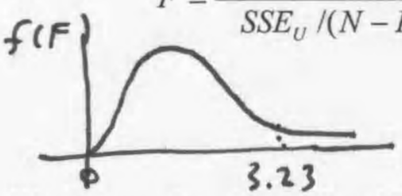
(3) SSE_R = sum of squared errors for the restricted model.

SSE_u = sum of squared errors for the unrestricted model.

(c) Using the above information and the following formula I want you to test the above null hypothesis versus the alternative hypothesis of $H_1: not H_0$ at the 5% significance level. Show your work below. Be sure that you indicate what the acceptance and rejection regions of your test are and what your conclusion is. → Reject $H_0: \beta_2 = \beta_3 = 0$. ceoten and ceotensq are significant variables.

(3)

$$F = \frac{(SSE_R - SSE_U)/J}{SSE_U/(N-K)} = \frac{(65-50)/2}{50/(44-4)} = \frac{7.5}{50} \cdot 40 = 7.5(.8) = 6.0$$



critical value $F_{2,44-4,.05} = 3.23$

$(0, 3.23)$ = acceptance region

$[3.23, \infty)$ = rejection region

4. ~~3~~. We can use the **additive/multiplicative dummy variable specification** of an equation to test that a basic economic relationship is different across groups (i.e. the **Chow Test**). Consider the Computer Output #1 that you have been given. Suppose we want to test whether the same regression model describes college grade point averages for male and female college athletes. The equation is

$$\text{Cumgpa} = \beta_0 + \beta_1 sat + \beta_2 hsperc + \beta_3 tothrs + u$$

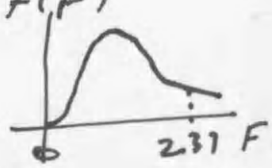
where **sat** is SAT score, **hsperc** is high school rank percentile, and **tothrs** is total hours of college courses.

(a) Use the computer output to test, at the $\alpha = 5\%$ level, the null hypothesis that there is no difference in this equation for male versus female athletes versus the alternative hypothesis that there is. Show your work below. Be sure to give me the acceptance and rejection regions of your test and the outcome of the test.

(3)

$$F = \frac{(SSE_R - SSE_u)/J}{SSE_u/(N-K)} = \frac{(547.36490 - 534.30915)/4}{534.30915/(732-8)} = 4.42$$

critical value = $F_{4,\infty,.05} = 2.37$



Acceptance region: $(0, 2.37)$

Rejection region: $[2.37, \infty)$

H_0: No difference between female and male equations

H_1: Difference between female and male equations.

Reject H_0. Accept H_1.

(b) Using Computer Output #1, I want you to write out the Cumgpa equation for the males (which is estimated using only male observation). Be sure to provide the coefficient estimates and below the estimates in parentheses I want you to include the standard errors of the estimates. Is hsperc an important variable in explaining cumgpa of **males**? Explain your answer.

(3)

Male equation: $\widehat{cumgpa} = \underset{(0.264)}{1.21398} + \underset{(0.000235)}{0.00061131}\, sat - \underset{(0.00178)}{0.00597}\, hsperc + \underset{(0.00109)}{0.01030}\, tothrs + \hat{e}_i$

The t-statistic for hsperc is $t = -3.36$ which is greater in abs. value than 2.0 thus is statistically significant. (Also $p = 0.0008 < 0.05$)

(c) Again using Computer Output # 1, is the effect of hsperc on cumgpa any different for males as compared to females: Explain your answer.

(2)

The t-statistic for the multiplicative dummy (female X hsperc) is 0.01 with $p = 0.9901 > 0.05$. Thus, it is not statistically significant and therefore we conclude that the male and female equations are not different with respect to how hsperc affects cumgpa.

5. A. Consider the results reported in Computer Output # 2.

The regression equation we are dealing with involves the price of a home (price) as a function of lot size (lotsize), square feet of floor space in the home (sqrft), and the number of bedrooms in the home (bdrms).

(a) State the null and alternative hypotheses of White's test for heteroskedasticity test.

H_0: Errors are homoskedastic
H_1: Errors are heteroskedastic

(3)

(b) Using a 5% level of significance, does it appear that the OLS regression is violating the assumption of homoskedasticity of the errors of the regression? Explain your answer. White's Heteroskedasticity Test with Cross-product terms produces an F-statistic with probability value of $p = 0.000010$ which is less than 0.05 therefore we conclude that the errors of the regression model are heteroskedastic.

(3)

(c) State at least two consequences for the Ordinary Least Squares procedure when heteroskedasticity is present in the errors of the regression model.

(2)

(1) OLS is no longer BLUE. It is not efficient relative to Weighted Least Squares.

(2) The OLS standard errors and t-statistics associated with the OLS estimates are invalid. Using them will result in often drawing incorrect statistical inferences.

(d) Given the results reported in Computer Output # 2, I want you to tell me what you think about the statement "Lotsize is a statistically significant variable in explaining the variation of house prices in the data that has been examined." Support your conclusion with statistical arguments. Given the results of White's heteroskedasticity test, one should use White's heteroskedasticity robust standard errors and t-statistics for the OLS coefficient estimates. Doing this we notice that the robust t-statistic on lotsize is 1.65 with p-value of $p = 0.1022 > 0.05$. Therefore, we conclude that lotsize is not a statistically significant explanatory variable.

(3)

(e) Suppose the regression equation in question has heteroskedasticity of the form

$$\sigma_i^2 = \sigma^2 lotsize \ .$$

In the below space I want you to write out the transformed house price regression equation that you would apply Ordinary Least Squares to in order to produce the Weighted Least Squares estimates of the coefficients of the house price equation.

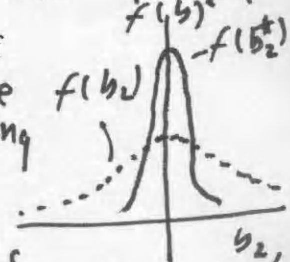
(3)

$$\frac{y_i}{\sqrt{lotsize}} = \frac{1}{\sqrt{lotsize_i}}\beta_1 + \beta_2\sqrt{lotsize_i} + \beta_3\frac{sqrft_i}{\sqrt{lotsize_i}} + \beta_4\frac{Bdrms_i}{\sqrt{lotsize_i}} + \frac{e_i}{\sqrt{lotsize_i}}$$

y_i = price

(f) Briefly to describe to me (in words and graphs) the major implication of **Aitken's theorem.** The WLS estimators are BLUE. The OLS estimators, though unbiased, have larger sampling variances than the corresponding WLS estimators.

(3)



---- = OLS sampling dist.
— = WLS sampling dist.

6. ~~5~~. Consider the following ARDL (ARX) model:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \beta_0 x_t + \beta_1 x_{t-1} + \cdots + \beta_r x_{t-r} + e_t \ .$$

(a)Write down the **autoregressive part** of the ARDL model.

(2) $\phi_1 y_{t-1} + \cdots + \phi_p y_{t-p}$

(b)Write down the **distributed lag part** of the ARDL model.

(2) $\beta_0 X_t + \beta_1 X_{t-1} + \cdots + \beta_r X_{t-r}$

(c) In order for the above regression equation to be **balanced**, what do we require? That both y_t and X_t be $I(0)$. That is y_t and X_t both need to be stationary time series.

(2)

(d) If the equation is **not** balanced, yet we go ahead apply OLS to the equation anyway, what is likely to be the result? Incorrect statistical inference.

(2)

(e) How do we know if the above equation is "**dynamically complete**?"

(2) There are enough lags in the autoregressive and distributed parts of the ARDL model to insure that the residuals of the estimated model are white noise (i.e. are uncorrelated at all lags.)

COMPUTER OUTPUT # 1

SAS CODE:

```
data gpa3;
  set gpa3;
  female_x_sat = female*sat;
  female_x_hsperc = female*hsperc;
  female_x_tothrs = female*tothrs;

proc reg data=gpa3;
  model cumgpa = female sat hsperc tothrs female_x_sat female_x_hsperc
female_x_tothrs;
  model cumgpa = sat hsperc tothrs;

run;
```

$$female = \begin{cases} 1 & \text{if female} \\ 0 & \text{otherwise} \end{cases}$$

The REG Procedure
Model: MODEL1
Dependent Variable: cumgpa

Number of Observations Read	732
Number of Observations Used	732

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	181.58941	25.94134	35.15	<.0001
Error	724	534.30915	0.73800		
Corrected Total	731	715.89856			

Root MSE	0.85907	R-Square	0.2537
Dependent Mean	2.08086	Adj R-Sq	0.2464
Coeff Var	41.28421		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	1.21398	0.26483	4.58	<.0001
female	1	-1.11364	0.52854	-2.11	0.0355
sat	1	0.00061131	0.00023503	2.60	0.0095
hsperc	1	-0.00597	0.00178	-3.36	0.0008
tothrs	1	0.01030	0.00109	9.43	<.0001
female_x_sat	1	0.00112	0.00050003	2.23	0.0258
female_x_hsperc	1	0.00005076	0.00410	0.01	0.9901
female_x_tothrs	1	0.00556	0.00207	2.69	0.0074

The REG Procedure
Model: MODEL2
Dependent Variable: cumgpa

Number of Observations Read	732
Number of Observations Used	732

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	168.53366	56.17789	74.72	<.0001
Error	728	547.36490	0.75187		
Corrected Total	731	715.89856			

Root MSE	0.86711	R-Square	0.2354
Dependent Mean	2.08086	Adj R-Sq	0.2323
Coeff Var	41.67060		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.92911	0.22855	4.07	<.0001
sat	1	0.00090283	0.00020787	4.34	<.0001
hsperc	1	-0.00638	0.00157	-4.07	<.0001
tothrs	1	0.01198	0.00093138	12.86	<.0001

Computer Output # 2

Dependent Variable: PRICE
Method: Least Squares
Date: 12/10/03 Time: 16:15
Sample: 1 88
Included observations: 88

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-21.77031	29.47504	-0.738601	0.4622
LOTSIZE	0.002068	0.000642	3.220096	0.0018
SQRFT	0.122778	0.013237	9.275093	0.0000
BDRMS	13.85252	9.010145	1.537436	0.1279

R-squared	0.672362	Mean dependent var	293.5460
Adjusted R-squared	0.660661	S.D. dependent var	102.7134
S.E. of regression	59.83348	Akaike info criterion	11.06540
Sum squared resid	300723.8	Schwarz criterion	11.17800
Log likelihood	-482.8775	F-statistic	57.46023
Durbin-Watson stat	2.109796	Prob(F-statistic)	0.000000

Dependent Variable: PRICE
Method: Least Squares
Date: 12/07/03 Time: 17:39
Sample: 1 88
Included observations: 88
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-21.77031	37.13821	-0.586197	0.5593
LOTSIZE	0.002068	0.001251	1.652283	0.1022
SQRFT	0.122778	0.017725	6.926707	0.0000
BDRMS	13.85252	8.478625	1.633817	0.1060

R-squared	0.672362	Mean dependent var	293.5460
Adjusted R-squared	0.660661	S.D. dependent var	102.7134
S.E. of regression	59.83348	Akaike info criterion	11.06540
Sum squared resid	300723.8	Schwarz criterion	11.17800
Log likelihood	-482.8775	F-statistic	57.46023
Durbin-Watson stat	2.109796	Prob(F-statistic)	0.000000

White Heteroskedasticity Test:

F-statistic	5.386953	Probability	0.000010
Obs*R-squared	33.73166	Probability	0.000100

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 04/20/06 Time: 20:51
Sample: 1 88
Included observations: 88

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	15626.24	11369.41	1.374411	0.1733
LOTSIZE	-1.859507	0.637097	-2.918719	0.0046
LOTSIZE^2	-4.98E-07	4.63E-06	-0.107498	0.9147
LOTSIZE*SQRFT	0.000457	0.000277	1.649673	0.1030
LOTSIZE*BDRMS	0.314647	0.252094	1.248135	0.2157
SQRFT	-2.673918	8.662183	-0.308689	0.7584
SQRFT^2	0.000352	0.001840	0.191484	0.8486
SQRFT*BDRMS	-1.020860	1.667154	-0.612337	0.5421
BDRMS	-1982.841	5438.483	-0.364595	0.7164
BDRMS^2	289.7541	758.8303	0.381843	0.7036

R-squared	0.383314	Mean dependent var	3417.316
Adjusted R-squared	0.312158	S.D. dependent var	7094.384
S.E. of regression	5883.814	Akaike info criterion	20.30444
Sum squared resid	2.70E+09	Schwarz criterion	20.58596
Log likelihood	-883.3955	F-statistic	5.386953
Durbin-Watson stat	2.052712	Prob(F-statistic)	0.000010

Table 3 **Right-Tail Critical Values for the *F*-Distribution**

Upper 5% Points

$v_2 \backslash v_1$	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	243.91	245.95	248.01	249.05	250.1	251.14	252.2	253.25	254.31
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.37
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	301	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.95	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

Source: This table was generated using the SAS® function FINV. v_1 = numerator degrees of freedom; v_2 = denominator degrees of freedom.