Exam 1 **Resource Economics 702**
March 10, 2005 **Econometrics I**

There are 100 points on this exam. You should spend no more that 1 minute per point. Thus, you will finish within 100 minutes (1 hour 40 minutes). Allocate your time accordingly. If you spend more than 15 minutes on each of these first two problems, you'll be over-budgeting your time.

Part I: On numerous occasions, I've stated that the disturbance makes life interesting in econometrics. ***In the following two theoretical questions, illustrate why this is true.***

(15) *1. We experimented with a known population regression function (PRF) early in this course. We appended a disturbance drawn from a known distribution to that PRF. Write an essay explaining how the disturbance played a role in those experiments. What was the main lesson from those experiments?*

a. Explain briefly how we modified the PRF to create a statistical relationship. How did this affect the results of individuals' estimations?

The known population regression function, our economic model, was $E[Y|X] = 4.0 + 1.0X$. To create the statistical model, we added a normally distributed disturbance with a mean of zero and variance of 16. Thus, we create the statistical model: $Y_i = \beta_0 + \beta_1 X_i + u_i$; where $\beta_0 = 4.0$ and $\beta_1 = 1.0$ were the known population parameters and $u \sim N(0,4)$ or $u \sim N(0,9)$.

Each person drew two random samples of disturbances from the same normal distribution for u and added those disturbance values to the same population regression function. Everyone had different random samples; we all started at different "seeds." Because each person added different disturbance values to the same population regression function, everyone had different sets of 40 Y_i-values for the same 40 X_i-values. We can say that *the X-values were fixed in repeated sampling*, one of the forms of CRM assumption # 2. ***Because the Y_i-values were different, each sample will yield different estimates of the two parameters,*** β_0 and β_1. One of our observations was that everyone had different estimates, b_0 and b_1, of the two population parameters β_0 and β_1. Another observation we should make is that all the randomness in this experiment is due to those different disturbance values. Everything else in the experiment was "controlled," everyone had exactly the same values.

b. What important statistical concept was our experiment designed to illustrate? Explain clearly what we observed about the estimators b_0 and b_1. *Your discussion should include the appropriate econometric jargon relating the population and its characteristics to samples, estimators and their characteristics.*

The different random samples, due to u, resulted in different estimates using the same estimators, the OLS estimators. The ***estimators are random variables***, one important concept that we illustrated by this experiment. Because b_0 and b_1 are random variables, they have distributions; we call their distributions ***sampling distributions***. The sampling distributions arise because each of us had different Y-values, random variables due to the disturbances we included. The variation in u was therefore responsible for the variation we observed in our values for b_0 and b_1. We investigated how this population variable, the disturbance, affected b_0 and b_1 by looking at histograms of our estimates. The histograms showed the centers of the sampling distributions were very close to the true values for β_0 and β_1. We also saw that different variances for u caused different variances for b_1. As $\text{var}(u)\uparrow$ so does $\text{var}(b_1)$. Thus, the sampling distribution for b_1 depends upon the distribution for u, which causes all the randomness/variation in our sample data and our estimators.

(15) *2. Ordinary Least Squares is the method by which we derived the OLS vector **b**. Doesn't this vector of OLS estimators satisfy our quest in econometrics? Please explain clearly why we needed the Classical Regression Model (CRM). In particular, you should:*

*a. explain clearly why and where the **6** CRM assumptions were important in developing results for **b**; and*

OLS satisfied one "quest" or goal in econometrics, ***point estimation***. But, we could go no further until we could describe the sampling distributions for b_0 and b_1. We had no real information about the sampling distributions, so we basically made a set of assumptions about the random variable that causes the sampling distributions to occur – the disturbance. ***The CRM assumptions were used to develop the expected values and variances of the estimators.*** To establish that the OLS estimators were unbiased, we used CRM assumptions 1 – 3. We then take this result and develop the variance by using CRM assumptions 4 and 5. The expected value of the estimator (center of the sampling distribution) and the variance (variation for the sampling distribution) are the two characteristics that are most important to us. (Having a linear estimator is important as is obvious from our work with expectations, but we don't get as jazzed up about it – we kind of take that property for granted.) Finally, the shape of the sampling distribution for **b**, or b_1, is provided by CRMA # 6. If the disturbances are normally distributed, then so is the random variable Y. Because we've shown that **b**, or b_1, is a linear combination of the Ys, then the OLS estimators are also normally distributed.

*b. **provide mathematical details** for your answer by focusing on our derivations of the two most important **desirable properties** of the OLS vector **b**. (I think it's easiest to do this in matrix notation, but if you prefer summation notation, just select one OLS estimator, say b_1. You can illustrate nicely using the simple linear model if you wish.)*

The first result we established was that the OLS estimators were unbiased. Using b_1 to illustrate, we begin with the estimator:

$$b_1 = \frac{\Sigma x_i Y_i}{\Sigma x_i^2} = \frac{\Sigma x_i(\beta_0 + \beta_1 X_i + u_i)}{\Sigma x_i^2}$$

We first needed to get the true value β_1 in this estimator ***- CRM Assumption 1 was used*** above to accomplish that task. We do a little algebra to simplify this expression:

$$b_1 = \frac{\Sigma x_i(\beta_0 + \beta_1 X_i + u_i)}{\Sigma x_i^2} = \beta_1 + \frac{\Sigma x_i u_i}{\Sigma x_i^2} .$$

To establish that b_1 is unbiased, we take the expected value,

$$E[b_1] = E\left[\beta_1 + \frac{\Sigma x_i u_i}{\Sigma x_i^2}\right],$$

and employ assumptions 2 and 3 of the CRM. Assumption 2 allow us to ignore Xs when taking expectations. Assumption 3 allows us to eliminate the final terms in the expression above, by CRM assumption #3 $E[u_i]$ is zero:

$$E[b_1] = \beta_1 + \frac{\Sigma x_i \ E[u_i]}{\Sigma x_i^2} = \beta_1$$

To derive the variance, we return to our proof that b_1 was unbiased where we found:

$$b_1 - \beta_1 = \frac{\Sigma x_i u_i}{\Sigma x_i^2} = \Sigma k_i u_i$$

Thus, we rely on CRM assumptions 1-3 to establish b_1 as an unbiased estimator and to find a starting point. Because b_1 is an unbiased estimator, $E[b_1] = \beta_1$ and the variance of b_1 is derived by (i) squaring the above result and (ii) taking the expectation:

$$E(b_1 - \beta_1)^2 = E[\Sigma k_i^2 u_i^2 + 2\sum_{i \neq j} k_i k_j u_i u_j]$$

When we take expectations, we find two kinds of terms. (Because $k_i = \frac{x_i}{\Sigma x_i^2}$, we needn't worry about expectations of the k_i - CRM assumption # 2.)

$$E[(b_1 - \beta_1)^2] = \Sigma k_i^2 \; E[u_i^2] + 2\sum_{i \neq j} k_i k_j \; E[u_i u_j]$$

Now, by CRM assumption 4 - $E[u_i^2] = \sigma^2$ and CRM assumption 5 - $E[u_i u_j] = 0$, giving us:

$$E[(b_1 - \beta_1)^2] = \sigma^2 \, \Sigma k_i^2 = \frac{\sigma^2}{\Sigma x_i^2} .$$

This variance is the minimum variance of all linear unbiased estimators by the Gauss-Markov Theorem.

By CRMA # 6, and the fact that the OLS estimators are linear combinations of Y (or u), we know that the OLS estimators will also be normally distributed.

Part II: Use the SAS output to answer the following questions.

To estimate U.S. demand for oil, time-series data were gathered for the period 1959 – 1999. Included in the data set are oil demand (million barrels), the price of oil ($/barrel), the price of coal ($ per short-ton), and gross national product ($ billions). A research assistant was asked to estimate the following general linear model:

$$qoil_t \;=\; \beta_0 \;+\; \beta_1 poil_t \;+\; \beta_2 pcoal_t \;+\; \beta_4 GNP_t + \; u_t$$

Unfortunately, he headed off to Florida for spring break leaving only his SAS program and some printouts. Please use the printouts to answer the following questions.

(10) a. He has a big asterisk on one set of results with the note: "Use these." Are these results consistent with the model I asked him to estimate? What did he do to generate these results and why do you think he did what he did?

These results (those marked by *) are not consistent with the model he was asked to estimate. He did the following: transformed all variables using the natural logarithm; estimated the model in "log-log" form.

There are two reasons he may have done this:
(1) he wants easy (elasticity) interpretations for all parameter estimates.
(2) he wishes to impose a "homogeneity" constraint. Theoretically, demand functions are homogeneous of degree zero.

(6) b. How should the parameter estimates be interpreted for the set of results my research assistant favors? Please write a sentence or two for each estimate providing proper interpretations.

These are elasticities:

b_1 = -0.36378; A 1% increase in the price of oil causes a 0.36378% decrease in oil demand, holding the price of coal and US GNP constant.

b_2 = 0.81634; A 1% increase in the price of coal causes a 0.816% increase in oil demand, holding the price of oil and GNP constant.

b_3 = 0.40567; A 1% increase in GNP causes a 0.406% increase in oil demand, holding the prices constant.

(10) c. Complete a hypothesis test for the effect of the price of oil on the U.S. demand for oil:

i. Specify both the null and alternative hypotheses;

$H_0: \; \beta_1 \geq 0$

$H_a: \; \beta_1 < 0$

ii. Choose the level of significance.

$\alpha = 0.10$ or $\alpha = 0.05$

iii. Identify the critical value(s) and draw a picture of your test.

$t_{(0.10,n-k)} \cong -1.303$ $t_{(0/05,37)} \cong -1.684$

iv. Calculate the test statistic.

t_{calc} = -2.43

v. State your conclusion.

Reject $H_0: \beta_1 \geq 0.$

There is sufficient evidence to conclude that the estimate is statistically less than zero.

(6) d. Comment briefly on the statistical significance of the remaining parameter estimates. Justify your conclusion in each case.

Both the remaining parameter estimates are statistically greater than zero. Both have positive calculated t-values that exceed the critical value of 1.303 (for a 10% level of significance).

(6) e. How well does the model "fit the data?" Explain.

The model explains 87.1% of the variation in oil demand. The R^2 statistic is used.

(8) f. Does the model explain a statistically significant portion of the variation in U.S. oil demand? Complete a hypothesis test. State both the null and alternative hypothesis, illustrate the test, calculate the appropriate test statistic and provide your conclusion.

This is an F-test; the appropriate F statistic is provided by the printout.
H_0: $\beta_1 = \beta_2 = \beta_3 = 0$; H_a: at least one is not zero. $\alpha = 0.05$

(8) g. Apparently, there is also some restriction that should be imposed for demand functions. What is the theoretical restriction? Explain briefly. Is the restriction supported by these data? Explain. How can you reach a conclusion?

The restriction is one of zero degree homogeneity. If all prices and income were to increase by the same percentage, then demand should not be affected. There should be no "money illusion."

The F-test is used: $F_{(1,37)} \sim 4.08$ for α=0.05

F_{calc} = 71.16.

The restriction is rejected. These data do not support the hypothesis that demand is homogeneous of degree zero.

(10) h. I really wanted a model that allows ***changing elasticities over the time series***. Did he provide a printout that will allow me to compute changing elasticities? Explain and then illustrate by calculating the income elasticity of U.S. oil demand for the first and final time period of the series.

The GLM specified initially would give parameter estimate that are partial effects of X and Y.

For example, for the linear model,

$$\beta_3 = \frac{\partial \text{ q oil}}{\partial \text{ GNP}}$$

These are constant over time, but the elasticity is calculated using our estimate from the linear model as:

$$\eta = b_3 \frac{\text{GNP}}{\text{qoil}}$$

Now, with a constant value for b_3, the elasticity changes as *GNP* and *qoil* change.

$$\eta_{1959} = (0.16911)\frac{\text{GNP}_{1959}}{\text{qoil}_{1959}} = (0.16911)\frac{2041.2}{1780}$$

$$\eta_{1959} = 0.1939$$

$$\eta_{1999} = (0.16911)\left(\frac{37047}{10852}\right) = \underline{\underline{0.5773}}$$

(6) i. Is the income elasticity calculated from a parameter estimate that is statistically different from zero? Explain.

The calculated t-value is 7.23. This value would easily exceed any critical value. The probability of a t_{calc} this high is less than 0.0001.

```
data newoil; set metrics.oil;
  lnoild=log(Oil_D); lnoilP=log(P_Oil); lncoalp=log(P_Coal); lnGnp=log(GNP);
run;
proc reg data=newoil;
  model Oil_D = P_oil P_Coal Gnp ;
    test P_oil + P_Coal + GNP = 0;
  model lnoild = lnoilp lncoalp lngnp ;
    test lnoilp + lncoalp + lngnp = 0;
run;
proc means data=metrics.oil;
run;
proc print data=newoil;
 var year Oil_D P_Oil P_Coal GNP;
run;
```

The REG Procedure
Model: MODEL1
Dependent Variable: Oil_D Oil D

Number of Observations Read	41
Number of Observations Used	41

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	243897272	81299091	50.22	<.0001
Error	37	59897053	1618839		
Corrected Total	40	303794326			

Root MSE	1272.33615	R-Square	0.8028
Dependent Mean	5905.63415	Adj R-Sq	0.7869
Coeff Var	21.54445		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	1522.87618	495.68419	3.07	0.0040
P_Oil	P Oil	1	-99.74630	49.18262	-2.03	0.0498
P_Coal	P Coal	1	220.55011	66.26276	3.33	0.0020
GNP	GNP	1	0.16911	0.02338	7.23	<.0001

The REG Procedure
Model: MODEL1

Test 1 Results for Dependent Variable Oil_D

Source	DF	Mean Square	F Value	Pr > F
Numerator	1	22994894	14.20	0.0006
Denominator	37	1618839		

The REG Procedure
Model: MODEL2
Dependent Variable: lnoild

Number of Observations Read	41
Number of Observations Used	41

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	11.05969	3.68656	83.17	<.0001
Error	37	1.64007	0.04433		
Corrected Total	40	12.69976			

Root MSE	0.21054	R-Square	0.8709
Dependent Mean	8.54826	Adj R-Sq	0.8604
Coeff Var	2.46293		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	3.53259	0.59256	5.96	<.0001
lnoilP		1	-0.36378	0.14981	-2.43	0.0202
lncoalp		1	0.81634	0.24090	3.39	0.0017
lnGnp		1	0.40567	0.06973	5.82	<.0001

The REG Procedure
Model: MODEL2

Test 1 Results for Dependent Variable lnoild

Source	DF	Mean Square	F Value	Pr > F
Numerator	1	3.15433	71.16	<.0001
Denominator	37	0.04433		

The MEANS Procedure

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
Year	Year	41	1979.00	11.9791486	1959.00	1999.00
P_Oil	P Oil	41	13.6431220	10.6561073	1.8000000	35.6900000
GNP	GNP	41	13562.88	10824.85	2041.20	37047.30
P_Coal	P Coal	41	15.6424390	8.2573860	4.5500000	27.2500000
Oil_D	Oil D	41	5905.63	2755.88	1780.00	10852.00
Pop	Pop	41	225.2732439	27.4928258	177.8290000	272.6900000

Obs	Year	Oil_D	P_Oil	P_Coal	GNP
1	1959	1780	2.080	4.95	2041.2
2	1960	1815	1.900	4.83	2122.4
3	1961	1917	1.800	4.73	2197.0
4	1962	2082	1.800	4.62	2362.9
5	1963	2123	1.800	4.55	2492.9
6	1964	2259	1.800	4.60	2677.5
7	1965	2468	1.800	4.55	2902.0
8	1966	2573	1.800	4.62	3177.9
9	1967	2537	1.800	4.69	3358.1
10	1968	2840	1.800	4.75	3670.4
11	1969	3166	1.800	5.08	3965.8
12	1970	3419	1.800	6.34	4184.3
13	1971	3926	2.240	7.15	4545.0
14	1972	4741	2.480	7.72	4996.4
15	1973	6256	3.290	8.59	5593.0
16	1974	6112	11.580	15.82	6066.9
17	1975	6056	11.530	19.35	6593.6
18	1976	7313	12.380	19.56	7364.2
19	1977	8807	13.300	19.95	8208.2
20	1978	8363	13.600	21.86	9272.0
21	1979	8456	30.030	23.75	10397.2
22	1980	6909	35.690	24.65	11323.3
23	1981	5996	34.280	26.40	12664.2
24	1982	5113	31.760	27.25	13182.5
25	1983	5051	28.770	25.98	14287.2
26	1984	5437	28.060	25.61	15872.3
27	1985	5067	27.530	25.20	16953.5
28	1986	6224	14.378	23.79	17873.4
29	1987	6678	18.423	23.07	19024.6
30	1988	7402	14.957	22.07	20507.0
31	1989	8061	18.196	21.82	22037.9
32	1990	8018	23.807	21.76	23328.9
33	1991	7627	20.047	21.49	24043.5
34	1992	7888	19.368	21.03	25369.4
35	1993	8620	17.067	19.85	26666.8
36	1994	8996	15.977	19.41	28284.6
37	1995	8835	17.179	18.83	29683.6
38	1996	9478	20.805	18.50	31325.0
39	1997	10162	19.301	18.14	33301.7
40	1998	10708	13.112	17.67	35112.3
41	1999	10852	18.251	16.76	37047.3