UNIVERSITY OF VICTORIA **EXAMINATIONS, DECEMBER 2014**

ECONOMICS 545: ECONOMETRIC ANALYSIS

TO BE ANSWERED IN BOOKLETS **INSTRUCTOR: D. Giles**
DURATION: 3 HOURS

STUDENTS SHOULD COUNT THE NUMBER OF PAGES IN THIS EXAMINATION PAPER BEFORE BEGINNING TO WRITE, AND REPORT ANY DISCREPANCY IMMEDIATELY TO THE INVIGILATOR. THIS QUESTION PAPER HAS 11 PAGES.

This is a "closed book/closed notes" examination. A Formula Sheet and Statistical Tables are provided. Calculators may be used.

ANSWER ANY *FOUR* QUESTIONS.
ALL QUESTIONS ARE WORTH EQUAL MARKS.
NO MARKS WILL BE GIVEN FOR EXTRA QUESTIONS THAT ARE ATTEMPTED.

(*Total Marks = 100*)

Question 1: (Total marks = 25)
You are on a co-op work term, and one of your colleagues comes out with the following comments about some applied econometric modeling you are both undertaking. Based on what you have learned in this course, in each case critically discuss the comments:

(a) "What's wrong with you? Autocorrelation really isn't a 'problem' – after all, the OLS estimator is still unbiased in this case."

5 marks

(b) "Obviously, my estimated equation is better than yours – the R^2 *and* the F-statistic in the EViews output are *both* higher in value than those for your equation."

5 marks

(c) "I don't like the way you started with lots of regressors in your model, and then eliminated some of them by using t-tests. It makes more sense to start off with the main regressor and then add some more variables into the model as long as they are statistically significant."

5 marks

(d) "Hey – don't tell the boss, but I used White's heteroskedasticity-consistent standard errors for this model, even though I'm quite sure that the errors are actually homoskedastic."

5 marks

(e) "You should have used Instrumental Variables estimation, as I did – we were told that the committee wanted an unbiased estimate of that price elasticity."

5 marks

Question 2: **(Total marks = 25)**

Suppose that the true data-generating process is

$$y = X_1\beta_1 + X_2\beta_2 + \varepsilon \;\; ; \quad \varepsilon \sim N(0,\, \sigma^2\Omega) \qquad (2.1)$$

where Ω is a *known* matrix.

However, the model that we estimate by mistake, using *Generalized Least Squares*, is

$$y = X_1\beta_1 + v\,. \qquad (2.2)$$

(a) *Prove* that the GLS estimator of β_1 will be biased, unless $(X_1'\Omega^{-1}X_2) = 0$.

7 marks

(b) How do you think the variability of this estimator will compare with the variability of the estimator of β_1 when the GLS estimator is applied to equation (2.1)?

4 marks

(c) Now let's suppose that the true data-generating process is given by

$$y = X_1\beta_1 + \varepsilon \;\; ; \quad \varepsilon \sim N(0,\, \sigma^2\Omega) \;\; , \qquad (2.3)$$

but the model that we estimate by GLS is

$$y = X_1\beta_1 + X_2\beta_2 + u\,. \qquad (2.4)$$

Prove that the GLS estimators of both β_1 and β_2 are *unbiased* in this case.

8 marks

[**Hint:** Note that $X_1 = (X_1 \;\; X_2)\begin{pmatrix} I \\ 0 \end{pmatrix} = XS$, say.]

(d) What other properties would you expect these last GLS estimators of β_1 and β_2 to possess? (*Just discuss this - no formal proofs are needed*.)

6 marks

Question 3: **(Total marks = 25)**

The models estimated in this question "explain" the accident rate in the U.S. coal mining industry, over the period 1940 – 1965. An important event during this period was the introduction of the Federal Coal Mine Safety Act in 1952. This law was intended to improve safety in the coal mining industry. The data series are:

F	Fatal injuries *per* million man-hours worked
NF	Non-fatal injuries *per* million man-hours worked
NFP	Non-fatal, permanent disability injuries *per* million man-hours worked
PER_MECH	Percentage of output of coal that is mechanically loaded (*not by hand*)
AV_OUT	Average output of coal *per* mine (tons)
FED_DUM	Federal regulation dummy (= 1, for 1953 – 1965; = 0 otherwise)

(a) Discuss the regression output in **RESULTS 1**, and comment on the quality of the model.

5 marks

(b) Why has the Newey-West procedure been used to estimate the standard errors?

2 marks

(c) What do you conclude when you compare **RESULTS 1** and **RESULTS 2**?

2 marks

RESULTS 1

Dependent Variable: LOG(F+NF+NFP)
Method: Least Squares
Date: 12/02/08 Time: 10:15
Sample: 1940 1965
Included observations: 26
Newey-West HAC Standard Errors & Covariance (lag truncation=2)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.876701	0.277843	17.55201	0.0000
FED_DUM	-0.083706	0.040008	-2.092211	0.0487
LOG(AV_OUT)	0.092329	0.037993	2.430166	0.0241
LOG(PER_MECH)	-0.297762	0.059043	-5.043182	0.0001
@TREND	0.002965	0.001271	2.332992	0.0297

R-squared	0.916910	Mean dependent var	3.981042
Adjusted R-squared	0.901084	S.D. dependent var	0.121867
S.E. of regression	0.038328	Akaike info criterion	-3.514215
Sum squared resid	0.030850	Schwarz criterion	-3.272273
Log likelihood	50.68480	Hannan-Quinn criter.	-3.444545
F-statistic	57.93486	Durbin-Watson stat	1.055717
Prob(F-statistic)	0.000000		

RESULTS 2

Dependent Variable: LOG(F+NF+NFP)
Method: Least Squares
Date: 12/02/08 Time: 10:17
Sample: 1940 1965
Included observations: 26

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.876701	0.385887	12.63763	0.0000
FED_DUM	-0.083706	0.030492	-2.745189	0.0121
LOG(AV_OUT)	0.092329	0.048029	1.922367	0.0682
LOG(PER_MECH)	-0.297762	0.073884	-4.030155	0.0006
@TREND	0.002965	0.003130	0.947052	0.3544

R-squared	0.916910	Mean dependent var	3.981042
Adjusted R-squared	0.901084	S.D. dependent var	0.121867
S.E. of regression	0.038328	Akaike info criterion	-3.514215
Sum squared resid	0.030850	Schwarz criterion	-3.272273
Log likelihood	50.68480	Hannan-Quinn criter.	-3.444545
F-statistic	57.93486	Durbin-Watson stat	1.055717
Prob(F-statistic)	0.000000		

RESULTS 3

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	6.042389	Prob. F(1,20)	0.0232
Obs*R-squared	6.032554	Prob. Chi-Square(1)	0.0140

Test Equation:
Dependent Variable: RESID
Method: Least Squares
Date: 12/02/08 Time: 10:20
Sample: 1940 1965
Included observations: 26
Presample missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.124277	0.350190	-0.354886	0.7264
FED_DUM	0.009418	0.027648	0.340631	0.7369
LOG(AV_OUT)	0.021628	0.044017	0.491355	0.6285
LOG(PER_MECH)	0.008127	0.066429	0.122340	0.9039
@TREND	-0.000117	0.002812	-0.041539	0.9673
RESID(-1)	0.496092	0.201817	2.458127	0.0232

(d) Interpret **RESULTS 3** and **RESULTS 4**. Explain why these two sets of results will be unaltered, whether they come from the regression in **RESULTS 1** or the regression in **RESULTS 2**.

4 marks

RESULTS 4

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	7.224276	Prob. F(2,19)	0.0046
Obs*R-squared	11.23105	Prob. Chi-Square(2)	0.0036

Test Equation:
Dependent Variable: RESID
Method: Least Squares
Date: 12/02/08 Time: 10:21
Sample: 1940 1965
Included observations: 26
Presample missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.092512	0.309242	-0.299159	0.7681
FED_DUM	-0.014515	0.026092	-0.556306	0.5845
LOG(AV_OUT)	0.019397	0.038849	0.499296	0.6233
LOG(PER_MECH)	0.000865	0.058682	0.014742	0.9884
@TREND	0.001501	0.002558	0.586860	0.5642
RESID(-1)	0.741291	0.201746	3.674372	0.0016
RESID(-2)	-0.553282	0.213947	-2.586073	0.0181

RESULTS 5

Dependent Variable: LOG(F+NF+NFP)
Method: Least Squares
Date: 12/02/08 Time: 10:25
Sample (adjusted): 1942 1965
Included observations: 24 after adjustments
Convergence achieved after 19 iterations
Newey-West HAC Standard Errors & Covariance (lag truncation=2)
MA Backcast: 1940 1941

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.762312	0.505983	11.38835	0.0000
FED_DUM	-0.072555	0.026789	-2.708367	0.0155
LOG(AV_OUT)	0.080981	0.016371	4.946674	0.0001
LOG(PER_MECH)	-0.503184	0.114672	-4.388012	0.0005
@TREND	0.006051	0.001577	3.836125	0.0015
AR(1)	0.614656	0.160236	3.835943	0.0015
AR(2)	-0.405270	0.114484	-3.539978	0.0027
MA(2)	-0.994983	0.056722	-17.54149	0.0000

R-squared	0.971970	Mean dependent var	3.965455
Adjusted R-squared	0.959706	S.D. dependent var	0.113187
S.E. of regression	0.022720	Akaike info criterion	-4.469917
Sum squared resid	0.008259	Schwarz criterion	-4.077232
Log likelihood	61.63900	Hannan-Quinn criter.	-4.365737
F-statistic	79.25881	Durbin-Watson stat	2.544452
Prob(F-statistic)	0.000000		

(e) Explain what has been assumed about the error term in the model that is estimated in **RESULTS 5**.

2 marks

(f) Using **RESULTS 5**, test whether or not the introduction of the Federal Coal Mine Safety Act had a significant impact on safety in the coal mining industry.

5 marks

(g) Interpret the estimated value of the coefficient of LOG(AV_OUT) in **RESULTS 5**, and construct a 95% confidence interval for this coefficient.

5 marks

<u>Question 4</u>: (Total marks = 25)

Suppose that we wish to estimate the standard multiple regression model, with fixed regressors:

$$y = X\beta + \varepsilon \ ; \quad \varepsilon \sim N(0, \sigma^2 I)$$

subject to a set of '*J*' exact linear restrictions: $R\beta = q$; where rank(R) = J.

(a) If e^* is the Restricted Least Squares (RLS) residual vector, and 'e' is the OLS residual vector, *prove* that:

$$(e^{*\prime}e^*) = (e'e) + (Rb - q)'[R(X'X)^{-1}R']^{-1}(Rb - q)$$

where 'b' is the OLS estimator of β.

6 marks

(b) Explain why $(e^{*\prime}e^{*})$ cannot be less than $(e'e)$.

3 marks

(c) Under what condition would $(e^{*\prime}e^{*}) = (e'e)$?

2 marks

(d) Explain how you could actually apply the restricted least squares estimator, using just OLS, in the case where the model is: $y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + \varepsilon_i$, and the restrictions that we want to impose are: $\beta_1 + \beta_2 = 1$, and $\beta_3 = \beta_4$.

5 marks

(e) Briefly discuss the finite-sample and asymptotic properties of this restricted estimator.

9 marks

<u>Question 5</u>: (Total marks = 25)

Suppose that we have a standard linear multiple regression model, with k regressors:

$$y = X\beta + \varepsilon , \qquad (5.1)$$

where the error term and regressors satisfy *all of the usual assumptions*. In addition, we have some further information in the form of J *uncertain* restrictions on the parameters:

$$R\beta = q + \nu , \qquad (5.2)$$

where R is $(J \times k)$, with rank(R) = J (< k); q is $(J \times 1)$; the elements of both R and q are *known*; and ν is a random error that reflects the uncertainty associated with the restrictions. You may assume that $\nu \sim N[0, V]$, and ε and ν are uncorrelated.

(a) Re-arrange equation (5.2) and then "stack up" this equation below equation (5.1) in a form that enables you to estimate the coefficient vector, β.

2 marks

(b) Show that the estimator for β, obtained by applying OLS to "stacked" model, is

$$\hat{\beta} = [X'X + R'R]^{-1}[X'y + R'q]$$

3 marks

(c) Prove that this estimator is unbiased, under the stated assumptions.

3 marks

(d) Derive the expression for the covariance matrix of this estimator.

3 marks

(e) Suppose that $V = \sigma^2 I$. Prove that in this case the covariance matrix of $\hat{\beta}$ simplifies to become $\sigma^2[X'X + R'R]^{-1}$.

3 marks

(f) For the situation in part (e), prove that $\hat{\beta}$ is more efficient than the OLS estimator, applied just to the model in (5.1).

5 marks

(g) For the situation in part (e), explain how you would construct a test statistic and actually test if one of the elements of the β vector takes a particular value, (say) β_0.

6 marks

<u>Question 6</u>: **(Total marks = 25)**

In this question we consider models that explain the number of witnesses (WIT) for each of the 210 reported UFO (Unidentified Flying Object) sightings in Canada in 2001. The data were obtained from the website for the Canadian UFO Survey.

The following explanatory variables will be considered:

DUR = duration of the sighting
REL = assessed reliability of the report(s) – value increases with reliability
STR = measure of "strangeness" of the UFO – value increases as the object gets more weird!
DBC = 1, if the primary sighting is in B.C. (0, otherwise)
DON = 1, if the primary sighting is in Ontario (0, otherwise)
DPR = 1, if the primary sighting is in a prairie province (0, otherwise)
DOUT = 1, if the number of witnesses is an "outlier" (0, otherwise)
(**Note:** DOUT = 1 for only 2 observations – ones involving 10 and 17 witnesses respectively)

(a) Discuss and interpret the regression output in **RESULTS A** below. Do the signs of the estimated coefficients seem plausible?

5 marks

(b) In **RESULTS B**, what null hypothesis is being tested, and what is the alternative hypothesis? What assumptions must be satisfied for this test to be appropriate? What do you conclude from these test results?

4 marks

(c) In **RESULTS C**, what has been done to take account of the conclusion that you reached in part (b) above?

2 marks

(d) **RESULTS D** are based on the model in **RESULTS C**. What is being tested in **RESULTS D**? What are the null and alternative hypotheses? What do you conclude about alien activity over B.C.!?

5 marks

(e) In **RESULTS E**, the variable RES_SQ is the series of squared residuals from the model in **RESULTS C**. What does this graph suggest about the error term in that model?

3 marks

(f) What estimator has been used in **RESULTS F**? What is the series called WEIGHT?

6 marks

RESULTS A

Dependent Variable: LOG(WIT)
Method: Least Squares
Date: 12/08/08 Time: 15:04
Sample: 1 210
Included observations: 210

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.003905	0.172354	0.022657	0.9819
LOG(DUR)	0.053386	0.015437	3.458280	0.0007
DOUT	2.170577	0.364062	5.962113	0.0000
REL	0.046148	0.023467	1.966540	0.0506
STR	-0.019938	0.032974	-0.604676	0.5461

R-squared	0.199917	Mean dependent var	0.453677
Adjusted R-squared	0.184306	S.D. dependent var	0.561991
S.E. of regression	0.507567	Akaike info criterion	1.505145
Sum squared resid	52.81297	Schwarz criterion	1.584838
Log likelihood	-153.0403	Hannan-Quinn criter.	1.537362
F-statistic	12.80587	Durbin-Watson stat	1.833507
Prob(F-statistic)	0.000000		

RESULTS B

Heteroskedasticity Test: White

F-statistic	2.092319	Prob. F(7,202)	0.0459
Obs*R-squared	14.19692	Prob. Chi-Square(7)	0.0478
Scaled explained SS	19.05241	Prob. Chi-Square(7)	0.0080

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 12/08/08 Time: 12:47
Sample: 1 210
Included observations: 210
Collinear test regressors dropped from specification

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.610263	0.368739	1.655001	0.0995
LOG(DUR)	0.035012	0.068485	0.511243	0.6097
(LOG(DUR))^2	-0.002332	0.005440	-0.428726	0.6686
(LOG(DUR))*REL	0.001253	0.008088	0.154911	0.8770
(LOG(DUR))*DOUT	-0.152676	0.262378	-0.581893	0.5613
REL	-0.227852	0.109451	-2.081772	0.0386
REL^2	0.023652	0.008989	2.631370	0.0092
REL*DOUT	0.135531	0.338483	0.400407	0.6893

RESULTS C

Dependent Variable: LOG(WIT)
Method: Least Squares
Date: 12/08/08 Time: 14:56
Sample: 1 210
Included observations: 210
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.263475	0.095271	2.765517	0.0062
LOG(DUR)	0.035619	0.017555	2.028989	0.0438
DOUT	1.993781	0.098099	20.32410	0.0000
DBC*REL	0.109052	0.060985	1.788175	0.0752
DBC*LOG(DUR)	0.043620	0.031329	1.392302	0.1653
DBC	-0.740677	0.320432	-2.311497	0.0218

R-squared	0.214189	Mean dependent var	0.453677
Adjusted R-squared	0.194929	S.D. dependent var	0.561991
S.E. of regression	0.504251	Akaike info criterion	1.496670
Sum squared resid	51.87089	Schwarz criterion	1.592302
Log likelihood	-151.1504	Hannan-Quinn criter.	1.535331
F-statistic	11.12089	Durbin-Watson stat	1.864769
Prob(F-statistic)	0.000000		

RESULTS D

Wald Test:
Equation: EQ_MODEL_1

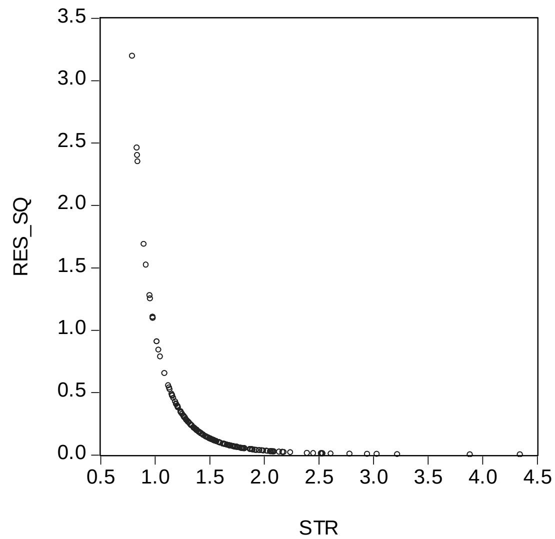
Test Statistic	Value	df	Probability
F-statistic	1.885655	(3, 204)	0.1331
Chi-square	5.656964	3	0.1295

Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
C(4)	0.109052	0.060985
C(5)	0.043620	0.031329
C(6)	-0.740677	0.320432

Restrictions are linear in coefficients.

RESULTS E



RESULTS F

Dependent Variable: LOG(WIT)
Method: Least Squares
Date: 12/08/08 Time: 15:29
Sample: 1 210
Included observations: 210
Weighting series: WEIGHT
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.143101	0.106419	-1.344699	0.1802
LOG(DUR)	0.065065	0.011047	5.889736	0.0000
DOUT	2.161144	0.065027	33.23457	0.0000
REL	0.046500	0.019307	2.408469	0.0169

Weighted Statistics

R-squared	0.348196	Mean dependent var	0.447666
Adjusted R-squared	0.338704	S.D. dependent var	0.542551
S.E. of regression	0.436868	Akaike info criterion	1.200494
Sum squared resid	39.31589	Schwarz criterion	1.264248
Log likelihood	-122.0518	Hannan-Quinn criter.	1.226267
F-statistic	36.68201	Durbin-Watson stat	1.827456
Prob(F-statistic)	0.000000		

Unweighted Statistics

R-squared	0.194946	Mean dependent var	0.453677
Adjusted R-squared	0.183222	S.D. dependent var	0.561991
S.E. of regression	0.507904	Sum squared resid	53.14112
Durbin-Watson stat	1.834877		

<u>Question 7</u>: **(Total marks = 25)**

Suppose that we have a linear multiple regression model that satisfies *all of the usual assumptions*, except that the errors follow a first-order autoregressive process. So:

$$y_t = \beta_1 + \beta_2 x_{2t} + ... + \beta_k x_{kt} + \varepsilon_t \qquad (7.1)$$
$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t \quad ; \quad u_t \sim i.i.d.\ N[0, \sigma_u^2] .$$

(a) Explain what implications this error structure has for: (i) the OLS estimator of the coefficient vector; (ii) the construction of confidence intervals for the regression coefficients; and (iii) the properties of the "t-statistics" associated with the estimated coefficients.

6 marks

(b) Multiply both sides of equation (7.1) by *ρ*, and lag everything by one period. Then subtract the equation you have just created from equation (7.1). Re-arrange the resulting equation so that the dependent variable is simply y_t, by itself.

3 marks

(c) Explain why it would make sense to estimate your new equation by Non-Linear Least Squares.

3 marks

(d) Discuss any practical issues that may arise when applying this estimator.

5 marks

(e) How could I use the Non-Linear Least Squares results to construct a simple test of the hypothesis that the errors in equation (7.1) are actually serially independent? What limitations might this test have?

5 marks

(f) What constraint must the parameter, *ρ*, in (7.1) satisfy? Why is this necessary, and what are the implications if this constraint is ***not*** satisfied?

3 marks

<u>END OF EXAMINATION</u>