Problem Set 8 Answer Key

Theoretical:

#1) given: $Y_i = \mu + u_i$ $\quad E[u_i/z_i] = 0 \quad var[u_i/z_i] = \sigma^2 z_i^2$

(a) In this particular heteroskedastic model in which the matrix "X" is a column of one's, the efficient estimator is the GLS estimator.

$Y_i = \mu X_i + u_i X_i \qquad var = \sigma^2 \Omega \qquad \therefore \Omega = \Sigma z_i^2$

$\Omega^{-1} = \Sigma \frac{1}{z_i^2}$

$$\hat{\beta}_{GLS} = (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} Y$$

$$\hat{\beta}_{GLS} = \left[\sum_{i=1}^{n} \frac{1}{z_i^2} X'X\right]^{-1} \left(\sum_{i=1}^{n} \frac{1}{z_i^2} X' y_i\right)$$

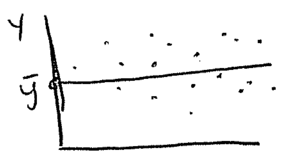
$$\hat{\beta}_{GLS} = \frac{\sum_{i=1}^{n} \frac{1}{z_i^2} y_i}{\sum_{i=1}^{n} \frac{1}{z_i^2}} = \boxed{\frac{\sum_{i=1}^{n} \frac{y_i}{z_i^2}}{\sum_{i=1}^{n} \frac{1}{z_i^2}}}$$

$$var\,\hat{\beta}_{GLS} = \sigma^2 (X' \Omega^{-1} X)^{-1}$$

$$= \sigma^2 \left(\sum_{i=1}^{n} \frac{1}{z_i^2} X'X\right)^{-1}$$

$$\boxed{= \frac{\sigma^2}{\sum_{i=1}^{n} \frac{1}{z_i^2}}}$$

(b) OLS estimator of β is $(X'X)^{-1} X'y$



$\boxed{b_{OLS} = \bar{y}}$

$$var(\bar{y}) = (X'X)^{-1} (X' \sigma^2 \Omega X)(X'X)^{-1}$$

where $(X'X) = \Sigma 1 \cdot 1 = n$

$(X'X)^{-1} = n^{-1}$

$$= n^{-1} \sigma^2 \sum_{i=1}^{n} z_i^2 \cdot n^{-1}$$

$$var(\bar{y}) = \boxed{\frac{\sigma^2 \sum_{i=1}^{n} z_i^2}{n^2}}$$

c) To show that the variance of the OLS estimator is greater than or equal to that of the GLS estimator, we must show that:

$$var(OLS) \geq var(GLS)$$

$$\left(\sigma^2/n^2\right)\sum_{i=1}^{n} z_i^2 \geq \sigma^2 / \sum_{i=1}^{n} \frac{1}{z_i^2}$$

$$\left(\frac{1}{n^2}\right)\left(\sum_i z_i^2\right)\left(\sum_s (1/z_s^2)\right) \geq 1$$

$$\sum_i \sum_s \left(z_i^2 / z_s^2\right) \geq n^2$$

The double sum contains n terms equal to one. Thus there remains $\left(\frac{n(n-1)}{2}\right)$ pairs of the form $\left(z_i^2/z_s^2 + z_s^2/z_i^2\right)$. If it can be shown that each of these sums is greater than or equal to 2, then we have proven our point.

Let $m_i = z_i^2$

must require: $m_i/m_s + m_s/m_i - 2 \geq 0$

equivalent to: $(m_i^2 + m_s^2 - 2m_i m_s)/m_i m_s \geq 0$

$= (m_i - m_s)^2 / m_i m_s \geq 0$

this must be true if m_i & m_s are positive, we know that they must be positive since $\boxed{m_i = z_i^2}$

#2) given: $Y_t = \beta' X_t + u_t$ where $\boxed{u_t = \rho u_{t-1} + \varepsilon_t}$

compare the autocorrelation in u_t in the original model to that of V_t in $Y_t - Y_{t-1} = \beta'(X_t - X_{t-1}) + V_t$ where $\boxed{V_t = u_t - u_{t-1}}$

find: $Cov[u_t, u_{t-1}]$ from original model

$Cov[\rho u_{t-1} + \varepsilon_t, u_{t-1}]$

$E[(\rho u_{t-1} + \varepsilon_t)(u_{t-1})]$

$E[\rho u_{t-1}^2 + \varepsilon_t u_{t-1}]$ ($E[\varepsilon_t u_{t-1}] = 0$)

$= \rho\, var[u_{t-1}]$

$\boxed{Cov[u_t, u_{t-1}] = \frac{\rho \sigma_\varepsilon^2}{1-\rho^2}}$

$Var(u_t) = \boxed{\frac{\sigma_\varepsilon^2}{1-\rho^2}}$

$$\text{autocorrelation coef.} = \boxed{\frac{Cov[u_t, u_{t-1}]}{Var(u_t)}}$$

$$\frac{\frac{\rho\sigma_\varepsilon^2}{1-\rho^2}}{\frac{\sigma_\varepsilon^2}{1-\rho^2}} \Rightarrow \boxed{\rho}$$

<u>new model</u>:

find: $Cov[V_t, V_{t-1}]$

$$V_t = u_t - u_{t-1}$$

$$Cov[\underbrace{u_t - u_{t-1}}_{V_t}, \underbrace{u_{t-1} - u_{t-2}}_{V_{t-1}}]$$

$$E[u_t u_{t-1} - u_t u_{t-2} - u_{t-1}^2 + u_{t-1}u_{t-2}]$$

→ use: $Cov[u_t, u_{t-s}] = \boxed{\frac{\rho^s\sigma_\varepsilon^2}{1-\rho^2}}$

$$= \left[\frac{\rho\sigma_\varepsilon^2}{1-\rho^2} - \frac{\rho^2\sigma_\varepsilon^2}{1-\rho^2} - \frac{\sigma_\varepsilon^2}{1-\rho^2} + \frac{\rho\sigma_\varepsilon^2}{1-\rho^2}\right] = \boxed{\frac{(2\rho - \rho^2 - 1)\sigma_\varepsilon^2}{1-\rho^2} = Cov[V_t, V_{t-1}]}$$

$$Var[V_t] = E[(u_t - u_{t-1})(u_t - u_{t-1})]$$

$$= E[u_t^2 - 2u_t u_{t-1} + u_{t-1}^2]$$

$$= \left[\frac{\sigma_\varepsilon^2}{1-\rho^2} - 2\frac{\rho\sigma_\varepsilon^2}{1-\rho^2} + \frac{\sigma_\varepsilon^2}{1-\rho^2}\right] = \boxed{\frac{(1-\rho)2\sigma_\varepsilon^2}{1-\rho^2} = Var[V_t]}$$

autocorrelation coef: $\frac{Cov(V_t, V_{t-1})}{Var(V_t)}$

$$= \frac{\frac{(2\rho - \rho^2 - 1)\sigma_\varepsilon^2}{1-\rho^2}}{\frac{2(1-\rho)\sigma_\varepsilon^2}{1-\rho^2}} \Rightarrow \frac{2\rho - \rho^2 - 1}{2(1-\rho)} \Rightarrow \frac{(\rho-1)(1-\rho)}{2(1-\rho)}$$

$$\Rightarrow \boxed{\frac{\rho-1}{2}}$$

Compare: $\boxed{\frac{\rho - 1}{2} \stackrel{?}{\ } \rho}$ * Plug in #'s for ρ to solve:

ρ	$\frac{\rho-1}{2}$
$-1/2$	$-3/4$
0	$-1/2$
$1/3$	$-1/3$
$1/2$	$-1/4$

* The autocorrelation of the differenced process reduces the absolute value of the autocorrelation coefficient when $\rho > 1/3$

$H_0: \rho = 0$ (no serial autocorrelation)
$H_1: \rho \neq 0$

(#3)

$\alpha = .05$
$T = 24$
$K = 2$ (# of indep. variables)
$k' = K - 1 = 2 - 1 = 1$

→ Careful not to use Durbin-Watson (Savin-White) tables

$DW_U = 1.446$
$DW_L = 1.273$
D.W. = 1.31

$1.273 < \underbrace{1.31}_{D.W.} < 1.446$

* Test is inconclusive

Empirical Problems

(#1) a) $\widehat{\ln earn}_w = 9.39 \Rightarrow \$12{,}076$ ($e^{\wedge}$)

b) $\widehat{\ln earn}_m = 9.88 \Rightarrow \$19{,}574$ ($e^{\wedge}$)

c) $\hat{\Delta}_{M-F} = \$7{,}498$

There are many reasons which could explain this difference

(#2) a) See SAS output
→ Coefs. are unbiased (β's)

b) The est. std. errors are biased because SAS uses $\sigma^2(X'X)^{-1}$ to calculate them, which is incorrect when heteroscedasticity is present.

c) we should use $\sigma^2(X'X)^{-1}X'\Omega X(X'X)$

d) use ACOV option in proc reg to get the new unbiased (consistent) std. errors, represented by the diagonal terms in the matrix which we then need to square root to get the actual std. error.
* Notice the variable Children is still inefficient, so the test didn't change anything.

e) no, estimates are inefficient, need to use FGLS to obtain efficient estimates.

#3

a) $H_0: \sigma_i = \sigma$ for all i

$H_A: \sigma_i \neq \sigma$

R^2 from whites regression = .0129

N = 1070

$NR^2 = (1070)(.0129) = 13.803$

$q = 19 - 1 = 18$

$$NR^2 \sim \chi^2_{(q)}$$

since 13.803 < 28.869

* we fail to reject the H_0

b) if $\Omega^{-1} = PP'$, is known then:

$$\hat{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1} X'\Omega^{-1}Y$$

if Ω is unknown then:

$$\hat{\beta}_{GLS} = (X'\hat{\Omega}^{-1}X)^{-1} X'\hat{\Omega}^{-1}Y$$

or $$\hat{\beta}_{GLS} = \left[\Sigma \frac{1}{\hat{e}_i^2} X_i X_i'\right]^{-1} \left[\Sigma \frac{1}{\hat{e}_i^2} X_i Y_i\right]$$

#4

a) estimates are unbiased

b) FGLS:

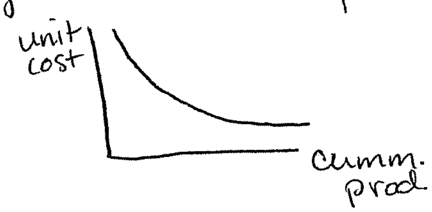
1. use OLS to get residuals $[e_i]$
2. regress e_i^2 on the variables to get $[\hat{e}_i^2]$
3. create new weights $[1/\hat{e}_i^2]$
4. use the weight in proc reg

c) all std. errors are lower (except for age).

d) consistency & better efficiency

#5

a) The coef. on cummulative production should be negative b/c as production ↑, unit cost will ↓ (learning curve)



#5

b) see SAS output (plot using excel)

→ yes autocorrelation

c) $$d = \frac{\sum_{t=2}^{T} (e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2} = \frac{.0231556}{.0349026} = .663 = d$$

D.W.$_u$ = 1.371

D.W.$_L$ = 1.106

T = 16

α = .05

K = 2

k' = 1

$H_0: \rho = 0$

$H_A: \rho \neq 0$

$d < D.W._L < D.W._u$

* reject H_0, so there is autocorrelation

d) Cochrane-Orcutt method: (See SAS code + output)

$$r = \frac{\sum_{t=2}^{T} e_t \cdot e_{t-1}}{\sum_{t=2}^{T} e_t^2} = \hat{\rho} = \frac{.0163730}{.0265726} = .6162 = \hat{\rho}$$

→ create transformed variables:

$$Y_{*t} = Y_t - \hat{\rho} Y_{t-1}$$

$$X_{*t} = X_t - \hat{\rho} X_{t-1}$$

→ run regression (throw out 1st obs)

$$\alpha_* = \alpha(1 - \hat{\rho})$$

$$\alpha_{FGLS} = \frac{\hat{\alpha}_*}{1 - \hat{\rho}}$$

$$Y_{*t} = \alpha_* + \beta X_{*t} + \varepsilon_{*t}$$

⇒ $\ln C_t = 2.74490 + -0.53019 \ln N_t + \varepsilon_{*t}$

SAS Output for Econ 216 Problem Set 8

**********************************Question 1a**********************************
Regression Using observations of women only

The REG Procedure
Model: MODEL1
Dependent Variable: lnearn

Number of Observations Read	470
Number of Observations Used	470

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	125.66998	41.88999	30.52	<.0001
Error	466	639.62077	1.37258		
Corrected Total	469	765.29075			

Root MSE	1.17157	R-Square	0.1642
Dependent Mean	**9.39982**	Adj R-Sq	0.1588
Coeff Var	12.46375		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	6.34089	0.38906	16.30	<.0001
AGE	1	0.00315	0.00410	0.77	0.4420
HIGRADE	1	0.10336	0.02039	5.07	<.0001
HOUR89	1	0.04120	0.00527	7.82	<.0001

************************Question 1b**
Average values of variables by gender

-- FEMALE=0 ------------------------------------

The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
lnearn	600	**9.8820997**	1.2170029	0	12.6619013
AGE	600	39.0783333	12.7496321	17.0000000	83.0000000
HIGRADE	600	13.0533333	2.9797013	0	20.0000000
HOUR89	600	42.9466667	10.4075783	15.0000000	99.0000000

-- FEMALE=1 ------------------------------------

Variable	N	Mean	Std Dev	Minimum	Maximum
lnearn	470	**9.3998226**	1.2773997	0	12.1794309
AGE	470	38.7744681	13.4532352	16.0000000	84.0000000
HIGRADE	470	13.3063830	2.7058654	0	20.0000000
HOUR89	470	37.8978723	10.2855912	15.0000000	99.0000000

************************** **Question 2a** **
OLS regression with ACOV matrix

The REG Procedure
Model: MODEL1
Dependent Variable: lnearn

Number of Observations Read	1070
Number of Observations Used	1070

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F

Model	5	198.99576	39.79915	27.96	<.0001
Error	1064	1514.77128	1.42366		
Corrected Total	1069	1713.76704			

Root MSE	1.19317	R-Square	0.1161
Dependent Mean	9.67026	Adj R-Sq	0.1120
Coeff Var	12.33856		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	8.03552	0.22021	36.49	<.0001
FEMALE	1	-0.47794	0.07377	-6.48	<.0001
AGE	1	0.00849	0.00303	2.80	0.0051
HIGRADE	1	0.09663	0.01285	7.52	<.0001
married	1	0.34964	0.08306	4.21	<.0001
CHILDREN	**1**	**0.02753**	**0.03543**	**0.78**	**0.4373**

**************************** Question 2d ***

OLS regression with ACOV matrix

The REG Procedure
Model: MODEL1
Dependent Variable: lnearn

Consistent Covariance of Estimates

Variable	Intercept	FEMALE	AGE	HIGRADE	married	CHILDREN
Intercept	**0.0549499195**	-0.004893581	-0.000544393	-0.00226341	0.0010979763	-0.002034124
FEMALE	-0.004893581	**0.0057890808**	0.0000690667	-0.00009673	0.001542294	0.0002354156
AGE	-0.000544393	0.0000690667	**0.0000129067**	6.9784254E-6	-0.00008168	0.0000103401
HIGRADE	-0.00226341	-0.00009673	6.9784254E-6	**0.0001563675**	-0.000216321	0.0000705155
married	0.0010979763	0.001542294	-0.00008168	-0.000216321	**0.0067319349**	-0.000341018
CHILDREN	-0.002034124	0.0002354156	0.0000103401	0.0000705155	-0.000341018	**0.0007618379**

	OLS Uncorrected Std Errors			**OLS Corrected Std Errors**		
Variable	Estimate	Std Err	t-stat	Estimate	Std Err (SQRT)	t-stat (est/std err)
Intercept	8.03552	0.22021	36.49	8.03552	0.23441	34.28
Female	-0.47794	0.07377	-6.48	-0.47794	0.07609	-6.28
age	0.00849	0.00303	2.8	0.00849	0.00359	2.36
higrade	0.09663	0.01285	7.52	0.09663	0.01250	7.73
married	0.34964	0.08306	4.21	0.34964	0.08205	4.26
children	*0.02753*	*0.03543*	*0.78*	*0.02753*	*0.02760*	*1.00*

**************************** Question 3 ***

Regression Results for Whites Test

The REG Procedure
Model: MODEL1
Dependent Variable: e2

Number of Observations Read	1070
Number of Observations Used	1070

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	18	1103.64804	61.31378	0.76	0.7446
Error	1051	84390	80.29488		
Corrected Total	1069	85494			

Root MSE	8.96074	R-Square	0.0129
Dependent Mean	1.41567	Adj R-Sq	-0.0040
Coeff Var	632.96636		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	4.28524	6.51338	0.66	0.5107
FEMALE	1	-2.80752	3.49193	-0.80	0.4216
AGE	1	-0.13979	0.17125	-0.82	0.4145
HIGRADE	1	0.10839	0.64791	0.17	0.8672
married	1	-2.01641	3.97467	-0.51	0.6120
CHILDREN	1	-0.36666	2.03223	-0.18	0.8569
fa	1	0.06181	0.04745	1.30	0.1930
fh	1	-0.02887	0.20105	-0.14	0.8858
fm	1	1.43312	1.26639	1.13	0.2580
fc	1	0.23277	0.56509	0.41	0.6805
ah	1	0.00372	0.00768	0.48	0.6283
am	1	0.03787	0.04828	0.78	0.4330
ac	1	-0.00842	0.03068	-0.27	0.7838
a2	1	0.00114	0.00155	0.74	0.4621
h2	1	-0.01078	0.01851	-0.58	0.5605
hm	1	-0.06095	0.22906	-0.27	0.7902
hc	1	0.02346	0.11192	0.21	0.8340
mc	1	0.62152	0.66070	0.94	0.3471
c2	1	-0.07357	0.20607	-0.36	0.7212

******************************** **Question 4b**

FGLS Regression

The REG Procedure
Model: MODEL1
Dependent Variable: lnearn

Number of Observations Read	1070
Number of Observations Used	1070

Weight: wtvar

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	217.52293	43.50459	42.96	<.0001
Error	1064	1077.38060	1.01258		
Corrected Total	1069	1294.90353			

Root MSE	1.00627	R-Square	0.1680
Dependent Mean	9.68913	Adj R-Sq	0.1641
Coeff Var	10.38554		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	7.59618	0.20442	37.16	<.0001
FEMALE	1	-0.47920	0.06797	-7.05	<.0001
AGE	1	0.01565	0.00311	5.04	<.0001
HIGRADE	1	0.10661	0.01283	8.31	<.0001
married	1	0.40903	0.07867	5.20	<.0001
CHILDREN	1	0.01926	0.02733	0.70	0.4811

*************************** Question 5a**
Learning Curve

The REG Procedure
Model: MODEL1
Dependent Variable: lnC

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	0.28598	0.28598	114.71	<.0001
Error	14	0.03490	0.00249		
Corrected Total	15	0.32088			

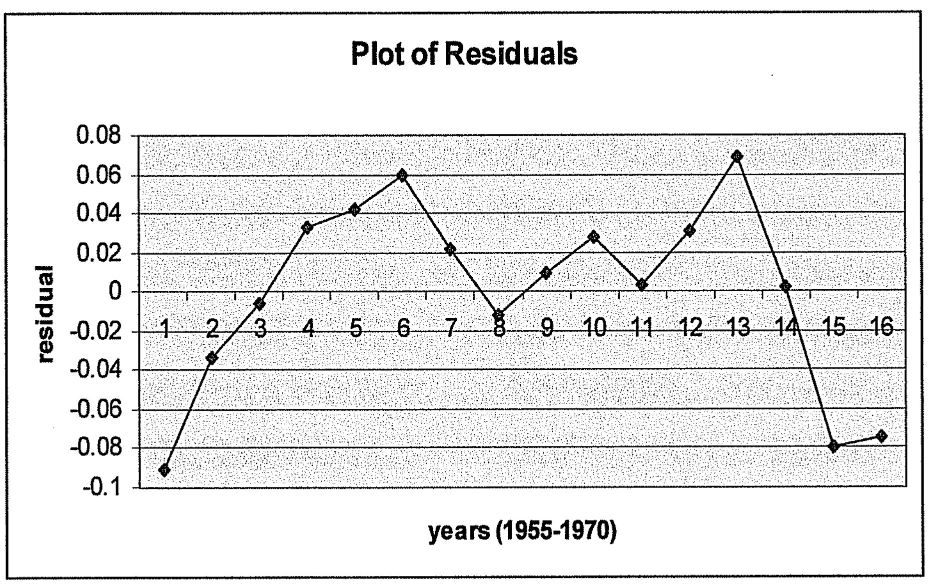
Root MSE	0.04993	R-Square	0.8912
Dependent Mean	3.07818	Adj R-Sq	0.8835
Coeff Var	1.62207		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	6.01909	0.27487	21.90	<.0001
lnN	1	-0.38570	0.03601	-10.71	<.0001

************************ Question 5b) Learning Curve **

obs	e
1	-0.091269
2	-0.033738
3	-0.005723
4	0.032655
5	0.041866
6	0.059178
7	0.02119
8	-0.011977
9	0.009295
10	0.028000
11	0.003567
12	0.031328
13	0.068329
14	0.001807
15	-0.079855
16	-0.074660



************************ Question 5c) **

The MEANS Procedure

```
Variable               Sum
ƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒ
edif2            0.0231550
e2               0.0349026
ƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒ
```

************************ Question 5d) **

The MEANS Procedure

```
Variable               Sum
ƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒ
e1e              0.0163730
e2               0.0265726
ƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒ
```

The REG Procedure
Model: MODEL1
Dependent Variable: lnct

Number of Observations Read	16
Number of Observations Used	15
Number of Observations with Missing Values	1

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.05884	0.05884	58.73	<.0001
Error	13	0.01302	0.00100		
Corrected Total	14	0.07186			

Root MSE	0.03165	R-Square	0.8188
Dependent Mean	1.16062	Adj R-Sq	0.8048
Coeff Var	2.72714		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	2.74490	0.20688	13.27	<.0001
lnnt	1	**-0.53019**	0.06918	-7.66	<.0001

SAS Code for Econ 216 Problem Set 8

```
libname j 'j:\classes\216';
libname g 'g:\';
run;

**************************** Problem 1 *********************************;

data census;
	set j.census90;
	if earn >0 then lnearn= log(earn);
	else lnearn =0;
	if marital=0 then married=1;
	else married=0;
run;

*a)restrict earnings equation to women;
Proc reg data=census;
	model lnearn= age higrade hour89;
	where female=1;
	title 'Regression Using observations of women only';
run;

*b)sort the data by gender in order to find the mean earnings of both men and women;
proc sort data=census;
	by female;
run;

proc means data=census;
	var lnearn age higrade hour89;
	by female;
	title 'Average values of variables by gender';
run;

*************************** Question 2 *********************************;

*runs the OLS regression for part a) and also adds ACOV matrix for part d);
proc reg data=census;
	model lnearn= female age higrade married children/ACOV;
	title 'OLS regression with ACOV matrix';
run;

************************* Question 3************************************;

*Whites General Test: in running the first regression I titled the output "whites1" and
the residual "e";
proc reg data=census noprint;
	model lnearn= female age higrade married children;
	output out=whites1 residual=e;
run;

*create a new data step and create interaction variables for every term.
note: dummy terms m2=married*married and f2=female*female are not included;
data whites2;
	set whites1;
	e2=e*e; fa=female*age; fh=female*higrade; fm=female*married;
	fc=female*children; ah=age*higrade; am=age*married;
	ac=age*children; a2=age*age; h2=higrade*higrade; hm=higrade*married;
	hc=higrade*children;
	mc=married*children; c2=children*children;
run;

*regress all variables on e2;
proc reg data=whites2;
	model e2=female age higrade married children fa fh fm fc ah am ac
	a2 h2 hm hc mc c2;
	title 'Regression Results for Whites Test';
run;
```

```
********************* Question 4b *******************************;

*starting off with same approach as in Q3;
proc reg data=census noprint;
       model lnearn=female age higrade married children;
       output out=fgls1 r=e;
run;

data fgls2;
       set fgls1;
       e2=e*e;
run;

*regress variables on e2 to get estimator e2hat;
proc reg data=fgls2 noprint;
       model e2= female age higrade married children;
       output out=fgls3 p=e2hat;
run;

*create new data step to get the weight 1/e2hat;
data fgls4;
       set fgls3;
       if e2hat <0 then delete;
       wtvar= 1/e2hat;
run;

*regress the model with the weights;
proc reg data=fgls4;
       model lnearn= female age higrade married children;
       weight wtvar;
       title 'FGLS Regression';
run;

******************* Question 5*********************************

*have to create C, which is the deflated unit cost;
data learn1;
       set j.tio2;
       c=ucostt*100/defl;
       if c>0 then lnC= log(c);
       else lnC=0;
       if cumt>0 then lnN= log(cumt);
       else lnN=0;
run;

*a)Learning curve;
proc reg data=learn1;
       model lnC=lnN;
       output out=learn2 r=e;
       title 'Learning Curve';
run;

*b)Plot the residuals;
proc print data=learn2;
       var e;
run;


*c) Durbin-Watson Test;
proc reg data=learn1 noprint;
       model lnC=lnN;
       output out=learn3 r=e;
run;

data dw;
       set learn3;
       le= lag(e);
       e2= e**2;
       edif2= (e-le)**2;
run;
```

```
proc means data=dw sum;
       var edif2 e2;
run;

*d) Cochrane-Orcutt Method;
*run OLS to get residual;
proc reg data=learn1 noprint;
       model lnC=lnN;
       output out=learn4 r=e;
run;

data CO;
       set learn4;
       le= lag(e);
       ele= e*le;
       e2= e**2;
run;

*to calculate row hat (ele/e2);
proc means data=CO sum;
       var ele e2;
       where yeart >1955;
run;

* to create transformed variables;
data transform;
       set learn1;
       llnc= lag (lnC);
       llnn= lag (lnN);
       lnct= lnC - 0.6162*llnc;
       lnnt= lnN - 0.6162*llnn;
run;

proc reg data=transform;
       model lnct= lnnt;
run;
```