$y_t = \alpha_1 + \alpha_2 \ln(x_t) + \alpha_3 k_t + \alpha_4 a_t + e_t$

y_t = weekly expenditure on transport.

x_t = weekly income.

k_t = number of children.

a_t = number of adults.

$\ln(x_t) = \ln x = \log(x_t)$.

A) $\boldsymbol{\delta}y/\boldsymbol{\delta}a = \alpha_4 = \alpha_t = 10$
B) $\boldsymbol{\delta}y/\boldsymbol{\delta}k = \alpha_3 = \alpha_k = 2$
C) $\boldsymbol{\delta}y/\boldsymbol{\delta}(\ln x) = (\boldsymbol{\delta}y/\boldsymbol{\delta}\ln x) * (\boldsymbol{\delta}\ln x/\boldsymbol{\delta}x) = \alpha_{\ln x} * 1/x = 0.03$ (from 3 cent).
Since $x = 800$; $\alpha_{\ln x} * 1/800 = 0.03$;
$\rightarrow \boldsymbol{\delta}y/(\ln x) = \alpha_2 = \alpha_{\ln(x)} = 24$.

Three alternative variance assumptions:

(I) $Var(e_t) = \sigma^2$ (Homoscedastic) (Run regression without transforming variables).
(II) $Var(e_t) = \sigma^2 \ln(x_t)^2$ (Heteroscedastic) (Transform data by dividing all variables by lnx).
(III) $Var(e_t) = \sigma^2 \ln(x_t)^4$ (Heteroscedastic) (Transform data by dividing all variables by $\ln x^2$).

(a).

Null and Alternative Hypothesis under three different conditions:

(A) $H_o: \alpha_a = 10$; $H_a: \alpha_a \neq 10$.

(B) $H_o: \alpha_k = 2$; $H_a: \alpha_k \neq 2$.

(c) $H_o: \alpha_{\ln(x)} = 24$; $H_a: \alpha_{\ln(x)} \neq 24$.

Data Management Process:

```
PROC IMPORT OUT= WORK.bm
            DATAFILE= "C:\Users\bmishra\Dropbox\Ph.D. Courseworks\Semest
er II, Spring 2019\Econometric Methods\Homeworks\Homework 4\HW4-DATA.xls"
            DBMS=EXCEL REPLACE;
     RANGE="hw9$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;

data bm; set bm;
/* Condition I */
y = trport;
a = a;
k = k;
x = x;
lnx = log(x);

/* White Test */
a2 = a**2;
k2 = k**2;
lnx2 = lnx**2;

/* Condition II */
ylnx = trport/lnx;
intlnx = 1/lnx;
alnx = a/lnx;
klnx = k/lnx;

/* Condition III */
ylnx2 = trport/lnx2;
intlnx2 = 1/lnx2;
alnx2 = a/lnx2;
klnx2 = k/lnx2;
run;

proc print;
run;
```

(b). Under Condition (I): Var(e_t) = σ^2

SAS Code:
```
proc reg data = bm;
model y = lnx a k;
run;
```

The REG Procedure
Model: MODEL1
Dependent Variable: y

Number of Observations Read	1000
Number of Observations Used	1000

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	604009	201336	26.21	<.0001
Error	996	7652096	7682.82710		
Corrected Total	999	8256105			

Root MSE	87.65174	R-Square	0.0732
Dependent Mean	79.41434	Adj R-Sq	0.0704
Coeff Var	110.37268		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	-127.86034	26.45058	-4.83	<.0001
lnx		1	28.35994	4.74543	5.98	<.0001
a	a	1	14.41463	5.29148	2.72	0.0066
k	k	1	0.71358	2.68037	0.27	0.7901

(b): Under Condition (II): $Var(e_t) = \sigma^2 \ln(x_t)^2$

```
proc reg data = bm;
model ylnx = intlnx alnx klnx; /* Condition II */
test alnx = 10;
test klnx = 2;
test intlnx = 24;
run;
```

The REG Procedure
Model: MODEL1
Dependent Variable: ylnx

Number of Observations Read	1000
Number of Observations Used	1000

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	6663.53464	2221.17821	11.82	<.0001
Error	996	187222	187.97420		
Corrected Total	999	193886			

Root MSE	13.71037	R-Square	0.0344
Dependent Mean	12.35960	Adj R-Sq	0.0315
Coeff Var	110.92886		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	22.67467	4.37849	5.18	<.0001
intlnx	1	-97.81402	23.33645	-4.19	<.0001
alnx	1	16.75278	5.14438	3.26	0.0012
klnx	1	1.91415	2.66594	0.72	0.4729

(b): Under condition (III): $Var(e_t) = \sigma^2 \ln(x_t)^4$

```
proc reg data = bm;
model ylnx2 = intlnx intlnx2 alnx2 klnx2 /noint; /* Condition III */
test alnx2 = 10;
test klnx2 = 2;
test intlnx = 24;
run;
```

The REG Procedure
Model: MODEL1
Dependent Variable: ylnx2

Number of Observations Read	1000
Number of Observations Used	1000

Note: No intercept in model. R-Square is redefined.

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	3859.82811	964.95703	196.95	<.0001
Error	996	4879.99931	4.89960		
Uncorrected Total	1000	8739.82742			

Root MSE	2.21350	R-Square	0.4416
Dependent Mean	1.94759	Adj R-Sq	0.4394
Coeff Var	113.65365		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
intlnx	1	14.01011	4.06656	3.45	0.0006
intlnx2	1	-54.66333	20.61977	-2.65	0.0082
alnx2	1	21.10203	5.03075	4.19	<.0001
klnx2	1	3.81634	2.69098	1.42	0.1564

Sensitivity Analysis:

Under first condition (homoscedasticity), the intercept is negative i.e. opposite to other model and variable related to weekly income is positive in first model and negative in other two models (heteroscedasticity). Other variables are consistently positive in all three models. In rest two cases corrected log (x) for heteroscedasticity has negative sign and other variables have positive sign. The significance of variables at 1% is consistent for all variables in all three models. Magnitude wise, the weekly income, number of children and number of adults are consistent in all three models. Income has highest coefficient and number of children has lowest.

(c). Condition I: Var(e_t) = σ^2 (Homoscedastic).

```
/*GQ Test Data Preparation */
data bm;
set bm;
proc sort;
by descending x;
run;
data bm1; set bm;/* Newly created data = bm1 and bm1 contains first 500
cases */
if x le 610;
run;
data bm2; set bm; /* Newly created data = bm1 and bm1 contains last 500
cases */
if x ge 611;
run;
proc reg data = bm1;
model y = lnx a k; /* Condition I */ /* Change this model for different
conditions */
output out = out1 r = ehat1;
run;
proc reg data = bm2;
model y = lnx4 k a; /* Change this model for different conditions */
output out = out2 r = ehat2;
run;
/* G-Q Test */
data bmout;
merge out1 out2;
keep ehat1 ehat2;
run;
proc means uss data = bmout;
var ehat1 ehat2;
output out = out3 uss = sse1 sse2;
run;
data bmout1; set out3;
x1 = 500; x2 = 500; k = 4;
sig1sq = sse1/(x1-k); sig2sq = sse2/(x2-k);
GQ = sig1sq/sig2sq;
run;
proc print;
run;
```

Null Hypothesis: Heteroscedasticity does not exist i.e. homoscedastic variance.

Alternative hypothesis: Heteroscedasticity does exist.

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: y

Number of Observations Read	500
Number of Observations Used	500

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	56072	18691	3.50	0.0154
Error	496	2647491	5337.68326		
Corrected Total	499	2703563			

Root MSE	73.05945	R-Square	0.0207
Dependent Mean	56.28340	Adj R-Sq	0.0148
Coeff Var	129.80639		

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	-19.04291	38.60112	-0.49	0.6220
lnx		1	9.11622	7.05265	1.29	0.1968
a	a	1	12.12005	6.16164	1.97	0.0497
k	k	1	3.48156	3.35385	1.04	0.2997

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: y

Number of Observations Read	500
Number of Observations Used	500

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	78601	26200	2.63	0.0495
Error	496	4938901	9957.46150		
Corrected Total	499	5017502			

Root MSE	99.78708	R-Square	0.0157
Dependent Mean	102.54528	Adj R-Sq	0.0097
Coeff Var	97.31026		

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	-140.53891	106.63194	-1.32	0.1881
lnx		1	31.06492	15.76378	1.97	0.0493
a	a	1	14.33199	8.74216	1.64	0.1018
k	k	1	-0.80998	4.14211	-0.20	0.8450

The MEANS Procedure

Variable	Label	Uncorrected SS
ehat1	Residual	2647490.90
ehat2	Residual	4938900.91

The SAS System

Obs	_TYPE_	_FREQ_	sse1	sse2	x1	x2	k	sig1sq	sig2sq	GQ
1	0	500	2647490.90	4938900.91	500	500	4	5337.68	9957.46	0.53605

GQ = 1/0.53605 = 1.8655.

F_{cric} (496, 496) $_{\alpha = 0.05}$ = 1.233 (Excel: =F.INV.RT(0.01,496, 496)]

Since GQ > F_{cric}, We reject null hypothesis that there is no heteroscedasticity. There might be problem of heteroscedasticity. This is due to weekly income.

(c). Condition II: $Var(e_t) = \sigma^2 \ln(x_t)^2$

```
/*GQ Test Data Preparation */
data bm;
set bm;
proc sort;
by descending x;
run;
data bm1; set bm;/* Newly created data = bm1 and bm1 contains first 500
cases */
if x le 610;
run;
data bm2; set bm; /* Newly created data = bm1 and bm1 contains last 500
cases */
if x ge 611;
run;
proc reg data = bm1;
model ylnx = intlnx alnx klnx; /* Condition II */ /* Change this model for
different conditions */
output out = out1 r = ehat1;
run;
proc reg data = bm2;
model ylnx = intlnx alnx klnx; /* Condition II */ /* Change this model for
different conditions */
output out = out2 r = ehat2;
run;
/* G-Q Test */
data bmout;
merge out1 out2;
keep ehat1 ehat2;
run;
proc means uss data = bmout;
var ehat1 ehat2;
output out = out3 uss = sse1 sse2;
run;
data bmout1; set out3;
x1 = 500; x2 = 500; k = 4;
sig1sq = sse1/(x1-k); sig2sq = sse2/(x2-k);
GQ = sig1sq/sig2sq;
run;
proc print;
run;
```

The REG Procedure
Model: MODEL1
Dependent Variable: ylnx

Number of Observations Read	500
Number of Observations Used	500

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	1456.12753	485.37584	3.01	0.0297
Error	496	79854	160.99661		
Corrected Total	499	81310			

Root MSE	12.68844	R-Square	0.0179
Dependent Mean	9.75234	Adj R-Sq	0.0120
Coeff Var	130.10664		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	2.85928	6.52467	0.44	0.6614
intlnx	1	9.48295	34.51935	0.27	0.7837
alnx	1	15.94301	6.16574	2.59	0.0100
klnx					

The REG Procedure
Model: MODEL1
Dependent Variable: ylnx

Number of Observations Read	500
Number of Observations Used	500

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	988.43977	329.47992	1.56	0.1983
Error	496	104789	211.26841		
Corrected Total	499	105778			

Root MSE	14.53508	R-Square	0.0093
Dependent Mean	14.96687	Adj R-Sq	0.0034
Coeff Var	97.11501		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	31.38960	16.27770	1.93	0.0544
intlnx	1	-144.44029	109.64818	-1.32	0.1883
alnx	1	14.87219	8.67768	1.71	0.0872
					0.9376

The MEANS Procedure

Variable	Label	Uncorrected SS
ehat1	Residual	79854.32
ehat2	Residual	104789.13

The SAS System

Obs	_TYPE_	_FREQ_	sse1	sse2	x1	x2	k	sig1sq	sig2sq	GQ
1	0	500	79854.32	104789.13	500	500	4	160.997	211.268	0.76205

GQ = 1/0.76205 = 1.312

F_{cric} (496, 496) $_{\alpha=0.05}$ = 1.233 (Excel: =F.INV.RT(0.01,496, 496)]

Since GQ > F_{cric}, We reject null hypothesis that there is no heteroscedasticity. There might be problem of heteroscedasticity. This is due to weekly income.

(c). Condition III: $Var(e_t) = \sigma^2 \ln(x_t)^4$

```
/*GQ Test Data Preparation */
data bm;
set bm;
proc sort;
by descending x;
run;
data bm1; set bm;/* Newly created data = bm1 and bm1 contains first 500
cases */
if x le 610;
run;
data bm2; set bm; /* Newly created data = bm1 and bm1 contains last 500
cases */
if x ge 611;
run;
proc reg data = bm1;
model ylnx2 = intlnx intlnx2 alnx2 klnx2 /noint; /* Condition III */ /*
Change this model for different conditions */
output out = out1 r = ehat1;
run;
proc reg data = bm2;
model ylnx2 = intlnx intlnx2 alnx2 klnx2 /noint; /* Condition III */ /*
Change this model for different conditions */
output out = out2 r = ehat2;
run;
/* G-Q Test */
data bmout;
merge out1 out2;
keep ehat1 ehat2;
run;
proc means uss data = bmout;
var ehat1 ehat2;
output out = out3 uss = sse1 sse2;
run;
data bmout1; set out3;
x1 = 500; x2 = 500; k = 4;
sig1sq = sse1/(x1-k); sig2sq = sse2/(x2-k);
GQ = sig1sq/sig2sq;
run;
proc print;
run;
```

The REG Procedure
Model: MODEL1
Dependent Variable: ylnx2

Number of Observations Read	500
Number of Observations Used	500

Note: No intercept in model. R-Square is redefined.

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	1562.53216	390.63304	76.62	<.0001
Error	496	2528.67500	5.09814		
Uncorrected Total	500	4091.20716			

Root MSE	2.25791	R-Square	0.3819
Dependent Mean	1.70692	Adj R-Sq	0.3769
Coeff Var	132.27964		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
intlnx	1	-6.31682	6.04975	-1.04	0.2969
intlnx2	1	50.25072	30.75843	1.63	0.1030
alnx2	1	21.73199	6.22417	3.49	0.0005
klnx2	1	6.99428	3.56959	1.96	0.0506

The REG Procedure
Model: MODEL1
Dependent Variable: ylnx2

Number of Observations Read	500
Number of Observations Used	500

Note: No intercept in model. R-Square is redefined.

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	2409.47234	602.36808	133.43	<.0001
Error	496	2239.14793	4.51441		
Uncorrected Total	500	4648.62026			

Root MSE	2.12471	R-Square	0.5183
Dependent Mean	2.18826	Adj R-Sq	0.5144
Coeff Var	97.09627		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
intlnx	1	31.37560	16.79470	1.87	0.0623
intlnx2	1	-146.04671	112.72038	-1.30	0.1957
alnx2	1	15.44292	8.61398	1.79	0.0736
klnx2	1	0.12109	4.07815	0.03	0.9763

The MEANS Procedure

Variable	Label	Uncorrected SS
ehat1	Residual	2528.68
ehat2	Residual	2239.15

The SAS System

Obs	_TYPE_	_FREQ_	sse1	sse2	x1	x2	k	sig1sq	sig2sq	GQ
1	0	500	2528.68	2239.15	500	500	4	5.09814	4.51441	1.12930

GQ = 1.12930

F_{cric} (496, 496) $_{\alpha = 0.05}$ = 1.233 (Excel: =F.INV.RT(0.01,496, 496)]

Since GQ < F_{cric}, We fail to reject null hypothesis. Thus, there is no heteroscedasticity i.e. homoscedasticity. The heteroscedasticity problem still appears in equations under first and second conditions.

(d): Under Condition (II): $Var(e_t) = \sigma^2 \ln(x_t)^2$

```
proc reg data = bm;
model ylnx = intlnx alnx klnx; /* Condition II */
test alnx = 10;
test klnx = 2;
test intlnx = 24;
run;
```

The REG Procedure
Model: MODEL1

Test 1 Results for Dependent Variable ylnx

Source	DF	Mean Square	F Value	Pr > F
Numerator	1	323.89011	1.72	0.1896
Denominator	996	187.97420		

Since p value is above 0.05, we fail to reject null hypothesis that adding an adult to a household increases household expenditure on transport by $10 per week.

The REG Procedure
Model: MODEL1

Test 2 Results for Dependent Variable ylnx

Source	DF	Mean Square	F Value	Pr > F
Numerator	1	0.19495	0.00	0.9743
Denominator	996	187.97420		

Since p value is above 0.05, we fail to reject null hypothesis that adding a child to a household increases household expenditure on transport by $2 per week.

The REG Procedure
Model: MODEL1

Test 3 Results for Dependent Variable ylnx

Source	DF	Mean Square	F Value	Pr > F
Numerator	1	5121.80865	27.25	<.0001
Denominator	996	187.97420		

Since p value is below 0.05, we reject null hypothesis that for a household with a weekly income of $800, an incremental increase in income increases household expenditure on transport by 3 cents per one dollar.

Under condition (III): $Var(e_t) = \sigma^2 \ln(x_t)^4$

```
proc reg data = bm;
model ylnx2 = intlnx intlnx2 alnx2 klnx2 /noint; /* Condition III */
test alnx2 = 10;
test klnx2 = 2;
test intlnx = 24;
run;
```

The REG Procedure
Model: MODEL1

Test 1 Results for Dependent Variable ylnx2

Source	DF	Mean Square	F Value	Pr > F
Numerator	1	23.86157	4.87	0.0276
Denominator	996	4.89960		

Since p value is below 0.05, we reject null hypothesis that adding an adult to a household increases household expenditure on transport by $10 per week.

The REG Procedure
Model: MODEL1

Test 2 Results for Dependent Variable ylnx2

Source	DF	Mean Square	F Value	Pr > F
Numerator	1	2.23221	0.46	0.4998
Denominator	996	4.89960		

Since p value is above 0.05, we fail to reject null hypothesis that adding a child to a household increases household expenditure on transport by $2 per week.

The REG Procedure
Model: MODEL1

Test 3 Results for Dependent Variable ylnx2

Source	DF	Mean Square	F Value	Pr > F
Numerator	1	29.56832	6.03	0.0142
Denominator	996	4.89960		

Since p value is below 0.05, we reject null hypothesis that for a household with a weekly income of $800, an incremental increase in income increases household expenditure on transport by 3 cents per one dollar.

(e).

```
/* White Heteroscedasticity Test */
proc reg data = bm;
model y = lnx a k; /* Condition I */
output out = white
p = whyhatt /* Predicted Value of dependent Variable y */
r = whyresid; /*Residual values of y */
run;

data white; set white;
whyressq = whyresid**2;
run;

proc reg data = white;
model whyressq = lnx k a lnx2 k2 a2;
test lnx = k = a = lnx2 = k2 = a2 = 0;
run;
```

The REG Procedure
Model: MODEL1
Dependent Variable: whyressq

Number of Observations Read	1000
Number of Observations Used	1000

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	8684393294	1447398882	3.21	0.0039
Error	993	4.472699E11	450422857		
Corrected Total	999	4.559543E11			

Root MSE	21223	R-Square	0.0190
Dependent Mean	7652.09580	Adj R-Sq	0.0131
Coeff Var	277.35105		

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	64006	41740	1.53	0.1255
lnx		1	-22846	13887	-1.65	0.1003
k	k	1	-2545.14494	1952.06276	-1.30	0.1926
a	a	1	658.66767	5959.47096	0.11	0.9120
lnx2		1	2156.63361	1130.88558	1.91	0.0568
k2		1	1019.96103	681.56325	1.50	0.1348
a2		1	-4.97433	1467.47957	-0.00	0.9973

Test 1 Results for Dependent Variable whyressq				
Source	DF	Mean Square	F Value	Pr > F
Numerator	6	1447398882	3.21	0.0039
Denominator	993	450422857		

Null Hypothesis: Homoscedasticity or No Heteroscedasticity.

Alternative Hypothesis: Heteroscedasticity.

Conclusion: Since p value is below 0.05, we reject the null hypothesis that there is no homoscedasticity.

(f).

```
/*Variance condition = Linear; Test = Lagrange Multiplier*/
proc autoreg data = bm;
model y = lnx k a;
hetero lnx / link = linear test = lm;
run;
```

The SAS System

The AUTOREG Procedure

Dependent Variable	y

The SAS System

The AUTOREG Procedure

Ordinary Least Squares Estimates			
SSE	7652095.8	DFE	996
MSE	7683	Root MSE	87.65174
SBC	11808.2429	AIC	11788.6119
MAE	60.0316125	AICC	11788.6521
MAPE	596.935203	HQC	11796.0731
Durbin-Watson	1.9444	Total R-Square	0.0732

Parameter Estimates

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > |t|	Variable Label
Intercept	1	-127.8603	26.4506	-4.83	<.0001	
lnx	1	28.3599	4.7454	5.98	<.0001	
k	1	0.7136	2.6804	0.27	0.7901	k
a	1	14.4146	5.2915	2.72	0.0066	a

ERROR: Convergence not attained in 50 iterations. Interpret the estimates with care.

The AUTOREG Procedure

Linear Heteroscedasticity Estimates			
SSE	7653037.9	Observations	1000
MSE	7653	Root MSE	87.48164
Log Likelihood	-5876.4972	Total R-Square	0.0730
SBC	11794.441	AIC	11764.9945
MAE	60.0205571	AICC	11765.0791
MAPE	595.002348	HQC	11776.1862
Hetero Test	56.7206	Normality Test	2166.3300
Pr > ChiSq	<.0001	Pr > ChiSq	<.0001

Parameter Estimates

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > |t|	Variable Label
Intercept	1	-124.6617	48.3539	-2.58	0.0099	
lnx	1	27.2750	7.8573	3.47	0.0005	
k	1	0.8077	2.4408	0.33	0.7407	k
a	1	16.2151	5.1102	3.17	0.0015	a
HET0	1	13.6501	102.8520	0.13	0.8944	
HET1	1	6.2324	96.2576	0.06	0.9484	

Null Hypothesis: Homoscedasticity.

Alternative Hypothesis: Heteroscedasticity.

Conclusion: Reject null hypothesis. (Looking at the p value)

```
/*Variance condition = Linear; Test = GLS */
proc autoreg data = bm;
model y = lnx k a /method = ml maxiter = 1000; /*    */
hetero lnx / link = linear test = lm; /*Variance condition = Linear */
run;
```

The AUTOREG Procedure

Ordinary Least Squares Estimates			
SSE	7652095.8	DFE	996
MSE	7683	Root MSE	87.65174
SBC	11808.2429	AIC	11788.6119
MAE	60.0316125	AICC	11788.6521
MAPE	596.935203	HQC	11796.0731
Durbin-Watson	1.9444	Total R-Square	0.0732

Parameter Estimates

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > |t|	Variable Label
Intercept	1	-127.8603	26.4506	-4.83	<.0001	
lnx	1	28.3599	4.7454	5.98	<.0001	
k	1	0.7136	2.6804	0.27	0.7901	k
a	1	14.4146	5.2915	2.72	0.0066	a

Algorithm converged.

The AUTOREG Procedure

Linear Heteroscedasticity Estimates			
SSE	7654497.23	Observations	1000
MSE	7654	Root MSE	87.48998
Log Likelihood	-5876.221	Total R-Square	0.0729
SBC	11793.8885	AIC	11764.442
MAE	60.094395	AICC	11764.5266
MAPE	599.673265	HQC	11775.6337
Hetero Test	56.7206	Normality Test	2175.5853
Pr > ChiSq	<.0001	Pr > ChiSq	<.0001

Parameter Estimates

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > |t|	Variable Label
Intercept	1	-113.5852	48.9430	-2.32	0.0203	
lnx	1	25.8140	7.9501	3.25	0.0012	
k	1	1.2568	2.4383	0.52	0.6062	k
a	1	15.0833	5.0978	2.96	0.0031	a
HET0	1	8.1348	175.3593	0.05	0.9630	
HET1	1	17.8081	774.4448	0.02	0.9817	

Null Hypothesis: Homoscedasticity.

Alternative Hypothesis: Heteroscedasticity.

Conclusion: Reject null hypothesis. (Looking at the p value).

Complete Code Compilation: Heteroscedasticity:

```
PROC IMPORT OUT= WORK.bm
            DATAFILE= "C:\Users\bmishra\Dropbox\Ph.D. Courseworks\Semest
er II, Spring 2019\Econometric Methods\Homeworks\Homework 4\HW4-DATA.xls"
            DBMS=EXCEL REPLACE;
     RANGE="hw9$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;

data bm; set bm;
/* Condition I */
y = trport;
a = a;
k = k;
x = x;
lnx = log(x);

/* White Test */
a2 = a**2;
k2 = k**2;
lnx2 = lnx**2;

/* Condition II */
ylnx = trport/lnx;
intlnx = 1/lnx;
alnx = a/lnx;
klnx = k/lnx;

/* Condition III */
ylnx2 = trport/lnx2;
intlnx2 = 1/lnx2;
alnx2 = a/lnx2;
klnx2 = k/lnx2;
run;
proc print;
run;

proc reg data = bm;
model y = lnx a k; /* Condition I */
test a = 10;
test k = 2;
test lnx = 24;
run;

proc reg data = bm;
model ylnx = intlnx alnx klnx; /* Condition II */
test alnx = 10;
test klnx = 2;
test intlnx = 24;
run;

proc reg data = bm;
```

```
model ylnx2 = intlnx intlnx2 alnx2 klnx2 /noint; /* Condition III */
test alnx2 = 10;
test klnx2 = 2;
test intlnx = 24;
run;

/*GQ Test Data Preparation */
data bm;
set bm;
proc sort;
by descending x;
run;
data bm1; set bm;/* Newly created data = bm1 and bm1 contains first 500
cases */
if x le 610;
run;
data bm2; set bm; /* Newly created data = bm1 and bm1 contains last 500
cases */
if x ge 611;
run;
proc reg data = bm1;
model ylnx2 = intlnx intlnx2 alnx2 klnx2 /noint; /* Condition III */ /*
Change this model for different conditions */
output out = out1 r = ehat1;
run;
proc reg data = bm2;
model ylnx2 = intlnx intlnx2 alnx2 klnx2 /noint; /* Condition III */ /*
Change this model for different conditions */
output out = out2 r = ehat2;
run;
/* G-Q Test */
data bmout;
merge out1 out2;
keep ehat1 ehat2;
run;
proc means uss data = bmout;
var ehat1 ehat2;
output out = out3 uss = sse1 sse2;
run;
data bmout1; set out3;
x1 = 500; x2 = 500; k = 4;
sig1sq = sse1/(x1-k); sig2sq = sse2/(x2-k);
GQ = sig1sq/sig2sq;
run;
proc print;
run;

/* E */ /* White Heteroscedasticity Test */
proc reg data = bm;
model y = lnx a k; /* Condition I */
output out = white
p = whyhatt /* Predicted Value of dependent Variable y */
r = whyresid; /*Residual values of y */
run;

data white; set white;
whyressq = whyresid**2;
```

```
run;

proc reg data = white;
model whyressq = lnx k a lnx2 k2 a2;
test lnx = k = a = lnx2 = k2 = a2 = 0;
run;
proc print;
run;

/* F */
/*Variance condition = Linear; Test = Lagrange Multiplier*/
proc autoreg data = bm;
model y = lnx k a;
hetero lnx / link = linear test = lm;
run;

/*Variance condition = Linear; Test = GLS */
proc autoreg data = bm;
model y = lnx k a /method = ml maxiter = 1000; /*   */
hetero lnx / link = linear test = lm; /*Variance condition = Linear */
run;
```

AGEC5213: *ECONOMETRIC METHODS*
Spring 2019

PROBLEM SET NO. 4 - due on April ~~18~~ 17, 2019

Question: Test three hypothesis listed in (A) (B) and (C) under three different conditions of variances (a) Homo. Var. (b) Hetero Var. and (c) Hetero Var.

Heteroskedasticity test and GLS

Oklahoma State is considering several policy changes in public transportation. Before approving any changes in transportation policy Governor Stitt wants to assess whether weekly expenditure figures provided by the Department of Transportation are realistic. For expenditure on transport, the Department believes that:

$\frac{dy}{dx}=10$; $H_0: \alpha_4 = 10$. $H_A: \alpha_4 \neq 10$

(A) Adding an adult to a household increases household expenditure on transport by $10 per week.

(B) Adding a child to a household increases household expenditure on transport by $2 per week. $\frac{dy}{d_0}=2$ $H_0: \alpha_3 = 2$ $H_A: \alpha_3 \neq 2$

(C) For a household with a weekly income of $800, an incremental increase in income increases household expenditure on transport by 3 cents per one dollar increase. $H_0: \alpha_2 = 24$ $H_A = \alpha_2 \neq 24$

$\frac{dy}{dx} = \frac{dy}{d\ln x} \cdot \frac{d\ln x}{dx} = \alpha_2 \cdot \frac{1}{x} = 0.03$ $X = 800$. $\alpha_2 = 24$

Your task is to provide the Governor helpful advice related to these issues. You have been retained to see if each of these rules of thumb are consistent with observed data. You decide to use data from the household expenditure survey that was conducted by the Department of Transportation of Oklahoma State. These data are stored in the file HW4-DATA.xls. You decide to set up the model.

$$y_t = \alpha_1 + \alpha_2 \ln(x_t) + \alpha_3 k_t + \alpha_4 a_t + e_t$$

where y_t represents weekly expenditure on transport, x_t represents weekly income and k_t and a_t represent the number of children and the number of adults in the household, respectively. Conscious of the fact that error terms in expenditure functions often have variances that depend on income, you decide to estimate the transport-expenditure function using three alternative variance assumptions, namely,

Homo Hetro Hetero

$$(i)\ \mathrm{var}(e_t) = \sigma^2 \quad (ii)\ \mathrm{var}(e_t) = \sigma^2[\ln(x_t)]^2 \quad (iii)\ \mathrm{var}(e_t) = \sigma^2[\ln(x_t)]^4$$

(a) Express each of the rules of thumb in terms of a null hypothesis that involves the parameters of the expenditure function. Specify the corresponding alternative hypotheses.

(b) Report the parameter estimates obtained under each error variance assumption and discuss the sensitivity of the estimates to this assumption (**Hint**: see GLS procedures in the lecture note). Run GLS

Sensitivity of the estimates: How sign, magnitude and significance of estimates differs in these three cases?

(c) For each of the equation estimates in part (b), use the Goldfeld-Quandt test to test for the existence of heteroskedasticity, where the heteroskedasticitiy is assumed to depend on x_t. Use a 1% level of significance with a one-tailed test for assumption that the variance of residuals increases with weekly income. In what equation(s) does heteroscedasticity still appear to be a problem based on the Goldfeld-Quandt test? Specify how you transform data

$H_0: \sigma_t^2 = \sigma^2$; $H_A: \sigma_t^2 = \sigma^2 x_t$ # Split data & test for each of three cases.

(d) Using each of the estimated equations under ii) and iii) and a 5% significance level, test the hypotheses specified in part (a). Comment on the results under the assumption that GLS has successfully removed heteroskedasticity. Question: Test above hypothesis using the regression equation assuming there is no heteroskedasticity in the model though model (ii) still has heteroskedasticity.

(e) Redo the White heteroskedasticity test at the 5% significance level for the equation estimates in

Include all variables & their permultations

N.

part (b) with

$(i)\ \text{var}(e_t) = \sigma^2$

(Hint: include independent variables and square of these variables in the variance equation). (No interaction)

(f) Using PROC AUTOREG procedure, test heteroskedasticity with LM test at the 5% level and find GLS and ML estimates. Your variance equation is assumed as: $\sigma_t^2 = \sigma^2(1+\gamma \ln x_t)$, where γ is the parameter of the variable $\ln x_t$.

$\boxed{\text{ln SAS}}$ hetero LNx

Question: The given condition of variance has is for linear variance as symbolized by the equation. Use method (1) under heading "How to do GLS/MLE in SAS" in sas code page. Other two conditions not mentioned here has non-linear variance.