AGEC5213: ***ECONOMETRIC METHODS***
Spring 2018

PROBLEM SET NO. 4 - due on April 2, 2018

Heteroskedasticity test and GLS (20 points)

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

rule of thumb

(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.

(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. ⇒ 주의 다시 check!!

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.

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

(A) $\frac{\partial y}{\partial a} = \alpha_4$

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,

(C) $\frac{\partial y}{\partial x} = \frac{\partial y}{\partial \ln x} \cdot \frac{\partial \ln x}{\partial x} = \alpha_2 \cdot \frac{1}{x}$

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

→ (A), (B), (C)

$\frac{\partial y}{\partial x} = \frac{\alpha_2}{x}$

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

(c) For each of the equation estimates in part (b), use the (1) Goldfeld-Quandt test to test for the existence of heteroskedasticity, where the heteroskedasticitiy is assumed to depend on x_t. (2) Use a 1% level of significance with a one-tailed test for assumption that the variance of residuals increases with transportation expense. In what equation(s) does heteroscedasticity still appear to be a problem based on the Goldfeld-Quandt test?

(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.

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

HW4 ECONOMETRICS METHODS Hyojae Jung

$Y_t = \alpha_1 + \alpha_2 \ln(d_t) + \alpha_3 k_t + \alpha_4 a_t + e_t$

three alternative variance assumptions

(i) $var(e_t) = \sigma^2$

(ii) $var(e_t) = \sigma^2 (\ln(d_t))^2$

(iii) $var(e_t) = \sigma^2 (\ln(d_t))^4$

(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 arternative hypotheses.

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

a_t : the number of adult $\quad Y_t$: weekly expenditure on transport.

$\Rightarrow \dfrac{\partial Y_t}{\partial a_t} = 10 = \alpha_4$

$H_0: \alpha_4 = 10 \quad H_a: \alpha_4 \neq 10$

(B) Adding a child to a household increases household expenditure on transport by \$2 per week.

k_t : the number of children. $\quad Y_t$: weekly expenditure on transport

$\Rightarrow \dfrac{\partial Y_t}{\partial k_t} = 2 = \alpha_3$

$H_0: \alpha_3 = 2 \quad H_a: \alpha_3 \neq 2$

(C) For a household with a weekly income of $800, [an increment in income increases household expenditure on transport by 3 per one dollar increase.]

x_t : weekly income $\qquad$ y_t : weekly expenditure on transport

$$\frac{\partial Y_t}{\partial x_t} = \underbrace{\frac{\partial Y_t}{\partial \ln x_t}}_{\alpha_2} * \frac{\partial \ln x_t}{\partial x_t} = \alpha_2 * \underbrace{\frac{\partial \ln x_t}{\partial x_t}}_{=\frac{1}{x_t}} = \alpha_2 * \frac{1}{x_t}$$

$$= \alpha_2 * \frac{1}{x_t} \text{ where } \boxed{x_t = \$800},$$

$$\Rightarrow \frac{\alpha_2}{800} = 0.03 \qquad \Rightarrow \alpha_2 = 24$$

$H_0 : \alpha_2 = 24 \qquad H_a : \alpha_2 \neq 24.$

(b) Report the parameter estimates obtained under each error variance assumption and the discuss the sensitivity of the estimatos to this assumption.

<Report the parameter estimates>

(i) $var(e_t) = \sigma^2$ → homoskedacity → just use OLS

(ii) $var(e_t) = \sigma^2[\ln(d_t)]^2$ → heteroscedacity → Data transformation → OLS

√ ← square root

$var(e_t) = \sigma^2[\ln(d_t)]^2$ ∴ We use $\sqrt{\ln(d_t)^2} = \ln d_t$

transformation data by using $\ln d_t$

$$\Rightarrow \frac{Y_t}{\ln(d_t)} = \frac{\alpha_1}{\ln(d_t)} + \alpha_2 + \alpha_3 \frac{k_t}{\ln(d_t)} + \alpha_4 * \frac{a_t}{\ln(d_t)} + \frac{e_t}{\ln(d_t)}$$

(iii) $var(e_t) = \sigma^2[\ln(d_t)]^4$ → heteroskedacity → Data transformation → OLS

$var(e_t) = \sigma^2[\ln(d_t)]^4$ ∴ We use $\sqrt{\ln(d_t)^4} = \ln(d_t)^2$

transformation data by using $\ln(d_t)^2$

$$\Rightarrow \frac{Y_t}{[\ln(d_t)]^2} = \frac{\alpha_1}{[\ln(d_t)]^2} + \frac{\alpha_2}{\ln(d_t)} + \frac{\alpha_3 * k_t}{[\ln(d_t)]^2} + \frac{\alpha_4 * a_t}{[\ln(d_t)]^2} + \frac{e_t}{[\ln(d_t)]^2}$$

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: trport trport

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
k	k	1	0.71358	2.68037	0.27	0.7901
a	a	1	14.41463	5.29148	2.72	0.0066

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: trportlnx

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
interlnx	1	-97.81402	23.33645	-4.19	<.0001
klnx	1	1.91415	2.66594	0.72	0.4729
alnx	1	16.75278	5.14438	3.26	0.0012

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: trportlnx2

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|
interlnx2	1	-54.66333	20.61977	-2.65	0.0082
interlnx	1	14.01011	4.06656	3.45	0.0006
klnx2	1	3.81634	2.69098	1.42	0.1564
alnx2	1	21.10203	5.03075	4.19	<.0001

−0.1

< sensitivity of estimates >

All estimates signs are same. All p-value have same result when $\alpha = 0.05$.

What about ≠ in magnitude?

From the results, estimates under different variance assumptions ~~are~~ vary. This means that estimates are sensitivity.

(check each estimate!)

(c) For each of the equation estimates in part (b), use the Goldfeld-Quandt test to test for the existence of heteroskedasticity, where the heteroskedacity, where the heteroskedasticity is assumed to depend on d_t. Use a 1% level of significance with one-tail test for assumption that the variance of residuals increases with transportation expense. In what equation(s) does heteroskedasticity still appear to be a problem based on the Goldfeld - Quandt test?

1 Step

We assume that the heteroskedasticity depends on d_t.

median of $d_t = 610.5$

We have to devide whole sample (total 1000) into two sub-sample based on $\boxed{d_t}$ ($n_1 > 610$, $n_2 \le 610$)

2. run separate regression for two groups

3. Compute estimated error variance $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$

(i) $var(e_t) = \sigma^2$

$SSE_1 = 2647491$ $SSE_2 = 4938935$

$$GQ_{(i)} = \frac{4938901 \leftarrow \text{larger}}{2647491 \leftarrow \text{smaller}} = 1.865502$$

(ii) $var(e_t) = \sigma^2 [\ln(d_t)]^2$

$SSE_1 = 79854$ $SSE_2 = 104789$

$$GQ_{(ii)} = \frac{104789 \leftarrow \text{larger}}{79854 \leftarrow \text{smaller}} = 1.312257$$

— 0.2

(iii) $var(e_t) = \sigma^2 [\ln(d_t)]^4$

$SSE_1 = 2528.675 \quad SSE_2 = 2239.14793$

$$GQ_{(3)} = \frac{2528.675 \leftarrow \text{larger}}{2239.14793 \leftarrow \text{smaller}} = 1.129302$$

$F_c = F_{496, 496, 0.01} \approx \underbrace{1.19}$ 1.232

$H_0: \underline{\sigma_t^2 = \sigma^2}$ homo , $H_a: \underline{\sigma_t^2 = \sigma^2 d_t}$ hetero

(i) $GQ_{(1)} = 1.865414 > F_{496, 496, 0.01} = 1.19$

⇒ We can reject null hypothesis at the significance level 0.01. We conclude that heteroskedasticity does exist

↳ CAUSED BY WEEKLY INCOME.

(ii) $GQ_{(2)} = 1.312227 > F_{496, 496, 0.01} = 1.19$

⇒ We can reject null hypothesis at the significance level 0.01. We conclude that heterockedasticity does exist

↳ SAME AS ABOVE

(iii) $GQ_{(3)} = 1.129295 < F_{496, 496, 0.01} = 1.19$

⇒ We cannot reject null hypothesis at the significance level 0.01. We conclude that heterockedasticity does not exist.

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: trport trport

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
k	k	1	3.48156	3.35385	1.04	0.2997
a	a	1	12.12005	6.16164	1.97	0.0497

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: trport trport

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
k	k	1	-0.80998	4.14211	-0.20	0.8450
a	a	1	14.33199	8.74216	1.64	0.1018

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: trportlnx

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
interlnx	1	9.48295	34.51935	0.27	0.7837
klnx	1	4.82041	3.43387	1.40	0.1610
alnx	1	15.94301	6.16574	2.59	0.0100

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: trportlnx

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
interlnx	1	-144.44029	109.64818	-1.32	0.1883
klnx	1	-0.32201	4.10886	-0.08	0.9376
alnx	1	14.87219	8.67768	1.71	0.0872

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: trportlnx2

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|
interlnx2	1	50.25072	30.75843	1.63	0.1030
interlnx	1	-6.31682	6.04975	-1.04	0.2969
klnx2	1	6.99428	3.56959	1.96	0.0506
alnx2	1	21.73199	6.22417	3.49	0.0005

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: trportlnx2

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|**
interlnx2	1	-146.04671	112.72038	-1.30	0.1957
interlnx	1	31.37560	16.79470	1.87	0.0623
klnx2	1	0.12109	4.07815	0.03	0.9763
alnx2	1	15.44292	8.61398	1.79	0.0736

(d) Using each of the estimated equations under (ii) iii) and a 5% significance level, test the hypothesis specified in part (a). Comment on the results under the assumption that GLS has successfully removed heteroskedasticity.

DON'T NEED TO REDO THIS.

(i) $var(e_t) = \sigma^2$

(A) $H_0 : \alpha_4 = 10$, $H_a : \alpha_4 \neq 10$

p-value 0.4043 > 0.05.

=> We cannot reject the null hypothesis at the signicance level 0.05

3 uks

(B) $H_0 : \alpha_3 = 2$, $H_a : \alpha_3 \neq 2$

p-value 0.6314 > 0.05

=> we cannot reject the null hypothesis at the significance level 0.05

(C) $H_0 : \alpha_2 = 24$ $H_a : \alpha_2 \neq 24$

p-value 0.3584 > 0.05

=> We cannot reject the null hypothesis at the significance level 0.05

-0.2

(iii) $var(e_t) = \sigma^2[\ln(t_i)]^4$

(A) H_0 $\alpha_4 = 10$ $H_a: \alpha_4 \neq 10$

p value $= 0.0276 < 0.05$

We can reject the null hypothesis at the significance level of 0.05
AND CONCLUDE?

(b) $H_0: \alpha_3 = 2$ $H_a: \alpha_3 \neq 2$

p-value $= 0.4998 > 0.05$

we cannot reject the null hypothesis at the significance level of 0.05
AND CONCLUDE?

(c) H_0. $\alpha_2 = 24$ $H_a: \alpha_2 \neq 24$

p-value $= 0.0142 < 0.05$

-0.2

we can reject the null hypothesis at the significance level of 0.05
CONCLUDE?

HOW DOES HK IMPACT INFERENCE? SHOULD STATE THAT UNDER HK WE ERRONEOUSLY FAIL TO REJECT H_0 FOR α_2 & α_4

In previous problem. Using Goldfeld-Quandt test ~~로 통해~~, we can identify in (iii) there is no heteroskedasticity.
~~However, In this part~~ ~~in II There are~~

(A), (C) → can reject
안 나오면 (B) → fail to reject vs. fail to reject
CI 때문에 CI 커지게 (즉 원래는 ~~확실히 hetero 해결~~ reject (CI가 커져
⇒ This means that if we estimate without considering hetero, the probability of erroneously concluding

(ii) $var(e_t) = \sigma^2[\ln(x_t)]^2$

(A) $H_0: \alpha_4 = 10$ $H_a: \alpha_4 \neq 10$

p-value $0.1896 > 0.05$

We cannot reject the null hypothesis at significance level of 0.05

(B) $H_0: \alpha_3 = 2$ $H_a: \alpha_3 \neq 2$

p-value $0.9743 > 0.05$

We cannot reject the null hypothesis at the significance level of 0.05

(C) $H_0: \alpha_2 = 24$ $H_a: \alpha_2 \neq 24$

p-value $0.7622 > 0.05$

We cannot reject the null hypothesis at the significance level of 0.05

−0.2

(iii) $var(e_t) = \sigma^2[\ln(t_i)]^4$

(A) $H_0: \alpha_4 = 10$ $H_a: \alpha_4 \neq 10$

p value = 0.0276 < 0.05

we can reject the null hypothesis at the significance level of 0.05

AND CONCLUDE?

(B) $H_0: \alpha_3 = 2$ $H_a: \alpha_3 \neq 2$

p-value = 0.4998 > 0.05

we cannot reject the null hypothesis at the significance level of 0.05

AND CONCLUDE?

(C) $H_0: \alpha_2 = 24$ $H_a: \alpha_2 \neq 24$

p-value = 0.0142 < 0.05

−0.2

we can reject the null hypothesis at the significance level of 0.05

CONCLUDE?

HOW DOES HK IMPACT INFERENCE? SHOULD STATE THAT UNDER HK WE ERRONEOUSLY FAIL TO REJECT H_0 FOR α_2 & α_4

In previous problem. Using Goldfeld-Quandt test ~~[illegible]~~, we can identify in (iii) there is no heteroskedasticity.

~~However, In this part~~ ~~in II There are~~

(A), (C) → can reject vs. fail to reject

(B) → fail to reject

~~왜냐하면~~, CI 계산이 CI → 커져서 (즉 원래는 ~~[illegible] hetero [illegible]~~ reject (CI가 커져

⇒ This means that if we estimate without considering hetero, the probability of erroneously concluding ↑

The SAS System

The REG Procedure
Model: MODEL1

Test 1 Results for Dependent Variable trport				
Source	**DF**	**Mean Square**	**F Value**	**Pr > F**
Numerator	1	5347.54977	0.70	0.4043
Denominator	996	7682.82710		

YOU CAN COPY AND PASTE THIS INTO ONE SINGLE PAGE.

The SAS System

The REG Procedure
Model: MODEL1

Test 2 Results for Dependent Variable trport				
Source	DF	Mean Square	F Value	Pr > F
Numerator	1	1769.68428	0.23	0.6314
Denominator	996	7682.82710		

The SAS System

The REG Procedure
Model: MODEL1

Test 3 Results for Dependent Variable trport				
Source	DF	Mean Square	F Value	Pr > F
Numerator	1	6485.32352	0.84	0.3584
Denominator	996	7682.82710		

The SAS System

The REG Procedure
Model: MODEL1

Test 1 Results for Dependent Variable trportlnx				
Source	DF	Mean Square	F Value	Pr > F
Numerator	1	323.89011	1.72	0.1896
Denominator	996	187.97420		

The SAS System

The REG Procedure
Model: MODEL1

Test 2 Results for Dependent Variable trportlnx				
Source	**DF**	**Mean Square**	**F Value**	**Pr > F**
Numerator	1	0.19495	0.00	0.9743
Denominator	996	187.97420		

The SAS System

The REG Procedure
Model: MODEL1

Test 3 Results for Dependent Variable trportlnx				
Source	DF	Mean Square	F Value	Pr > F
Numerator	1	17.22255	0.09	0.7622
Denominator	996	187.97420		

The SAS System

The REG Procedure
Model: MODEL1

Test 1 Results for Dependent Variable trportlnx2				
Source	DF	Mean Square	F Value	Pr > F
Numerator	1	23.86157	4.87	0.0276
Denominator	996	4.89960		

The SAS System

The REG Procedure
Model: MODEL1

Test 2 Results for Dependent Variable trportlnx2

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

The SAS System

The REG Procedure
Model: MODEL1

Test 3 Results for Dependent Variable trportlnx2				
Source	DF	Mean Square	F Value	Pr > F
Numerator	1	29.56832	6.03	0.0142
Denominator	996	4.89960		

(e) Redo the White heteroskedasticity test at the 5% significance level for the equation estimates in part(b) with (i) $var(e_t) = \sigma^2$

~~equation~~

$$\hat{e}_t^2 = \beta_0 + \beta_1 \ln(d_t) + \beta_2 k_t + \beta_3 a_t + \beta_4 (\ln d_t)^2 + \beta_5 k_t^2 + \beta_6 a_t^2 + \beta_7 \ln(d_t) k_t + \beta_8 \ln(d_t) a_t + \beta_9 k_t a_t + u_t$$

$H_0: \beta_1 = \beta_2 = \beta_3 = \cdots = \beta_9 = 0$ $\quad H_a$: At least one is not zero.

-0.2 State H_0 as: $VAR(e_t) = \sigma^2$ vs $H_A: VAR(e_t) \neq \sigma^2$

p-value = 0.0006

$\therefore$ p-value $= 0.0006 < 0.05 = \alpha$

We can reject the null hypothesis at the significance level of 0.05. It means that there is heteroscedasticity.

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: ehatsqr

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	9	13231236764	1470137418	3.29	0.0006
Error	990	4.427231E11	447195004		
Corrected Total	999	4.559543E11			

Root MSE	21147	R-Square	0.0290
Dependent Mean	7652.09580	Adj R-Sq	0.0202
Coeff Var	276.35548		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	103727	43758	2.37	0.0180
lnx		1	-47311	15837	-2.99	0.0029
k	k	1	11581	7968.18147	1.45	0.1464
a	a	1	33682	13127	2.57	0.0104
lnx2		1	5302.96821	1500.54809	3.53	0.0004
k2		1	929.82343	694.86507	1.34	0.1812
a2		1	2790.89023	1883.29449	1.48	0.1387
lnxa		1	-7057.95282	2576.43608	-2.74	0.0063
lnxk		1	-2391.15898	1327.87679	-1.80	0.0720
ak		1	867.64545	1512.75082	0.57	0.5664

The SAS System

The REG Procedure
Model: MODEL1

Test 1 Results for Dependent Variable ehatsqr				
Source	DF	Mean Square	F Value	Pr > F
Numerator	9	1470137418	3.29	0.0006
Denominator	990	447195004		

(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 : $\sigma_t^2 = \sigma^2(1+\gamma \ln d_t)$, where γ is the parameter of the variable $\ln d_t$

$H_0 : \gamma = 0$, $H_a : \gamma \neq 0$

-0.2

(i) GLS

each p-value = <0.0001 < 0.05

=> we can reject null hypothesis. It means that there is heteroscedasticity CAUSED BY WEEKLY INCOME

-0.2

(ii) MLE

each p-value = <0.0001 < 0.05

=> We can reject null hypothesis It means that there is heteroscedasticity. CAUSED BY WEEKLY INCOME

YOUR ML ESTIMATES SHOULD BE SLIGHTLY ≠ THAN GLS.
YOUR ALGORITHM DID NOT CONVERGE. CHECK YOUR MODEL.

$\boxed{19.4 / 20}$

The SAS System

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	

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	

The SAS System

The AUTOREG Procedure

Linear Heteroscedasticity Estimates

SSE	7653564.73	**Observations**	1000
MSE	7654	**Root MSE**	87.48465
Log Likelihood	-5873.6226	**Total R-Square**	0.0730
SBC	11802.5072	**AIC**	11763.2452
MAE	60.2121497	**AICC**	11763.3905
MAPE	597.148661	**HQC**	11778.1675
Hetero Test	68.9732	**Normality Test**	2284.6610
Pr > ChiSq	<.0001	**Pr > ChiSq**	<.0001

Parameter Estimates

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > |t|	Variable Label
Intercept	1	-127.2658	45.9551	-2.77	0.0056	
lnx	1	27.6736	7.4368	3.72	0.0002	
k	1	1.1731	4.3113	0.27	0.7855	k
a	1	16.3515	8.0250	2.04	0.0416	a
HET0	1	19.3632	70.8234	0.27	0.7845	
HET1	1	1.9602	15.3589	0.13	0.8984	
HET2	1	1.2520	9.1664	0.14	0.8914	
HET3	1	2.8189	21.0386	0.13	0.8934	

What ARE THESE?

The AUTOREG Procedure

Linear Heteroscedasticity Estimates

SSE	7653564.73	Observations	1000
MSE	7654	Root MSE	87.48465
Log Likelihood	-5873.6226	Total R-Square	0.0730
SBC	11802.5072	AIC	11763.2452
MAE	60.2121497	AICC	11763.3905
MAPE	597.148661	HQC	11778.1675
Hetero Test	68.9732	Normality Test	2284.6610
Pr > ChiSq	<.0001	Pr > ChiSq	<.0001

Parameter Estimates

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > |t|	Variable Label
Intercept	1	-127.2658	45.9551	-2.77	0.0056	
lnx	1	27.6736	7.4368	3.72	0.0002	
k	1	1.1731	4.3113	0.27	0.7855	k
a	1	16.3515	8.0250	2.04	0.0416	a
HET0	1	19.3632	70.8234	0.27	0.7845	
HET1	1	1.9602	15.3589	0.13	0.8984	
HET2	1	1.2520	9.1664	0.14	0.8914	
HET3	1	2.8189	21.0386	0.13	0.8934	

```
Data hw4; set hw4;
lnx= log(x);
lnx2 = lnx*lnx;
trportlnx = trport/lnx;
interlnx = 1/lnx;
klnx = k/lnx;
alnx = a/lnx;
trportlnx2 = trport/lnx2;
interlnx2 = 1/lnx2;
klnx2 = k/lnx2;
alnx2 = a/lnx2;
run;
proc reg data = hw4;
model trport = lnx k a;
test a=10;
test k=2;
test lnx=24;
output out=white r=ehat1;
run;
proc reg data = hw4;
model trportlnx = interlnx klnx alnx;
test alnx=10;
test klnx=2;
test intercept=24;
run;
proc reg data = hw4;
model trportlnx2 = interlnx2 interlnx klnx2 alnx2 / noint;
test alnx2=10;
test klnx2=2;
test interlnx=24;
run;
Data hw4gold1; set hw4;
if x>610 then delete;
run;
proc reg data=hw4gold1;
model trport = lnx k a;
run;
Data hw4gold2; set hw4;
if x<=610 then delete;
run;
proc reg data=hw4gold2;
model trport = lnx k a;
run;
proc reg data=hw4gold1;
model trportlnx = interlnx klnx alnx;
run;
proc reg data=hw4gold2;
model trportlnx = interlnx klnx alnx;
run;
proc reg data=hw4gold1;
model trportlnx2 = interlnx2 interlnx klnx2 alnx2 /noint;
run;
proc reg data=hw4gold2;
model trportlnx2 = interlnx2 interlnx klnx2 alnx2 /noint;
run;
Data whitehetero; set white;
ehatsqr = ehat1*ehat1;
```

```
Data hw4; set hw4;
lnx= log(x);
lnx2 = lnx*lnx;
trportlnx = trport/lnx;
interlnx = 1/lnx;
klnx = k/lnx;
alnx = a/lnx;
trportlnx2 = trport/lnx2;
interlnx2 = 1/lnx2;
klnx2 = k/lnx2;
alnx2 = a/lnx2;
run;
proc reg data = hw4;
model trport = lnx k a;
test a=10;
test k=2;
test lnx=24;
output out=white r=ehat1;
run;
proc reg data = hw4;
model trportlnx = interlnx klnx alnx;
test alnx=10;
test klnx=2;
test intercept=24;
run;
proc reg data = hw4;
model trportlnx2 = interlnx2 interlnx klnx2 alnx2 / noint;
test alnx2=10;
test klnx2=2;
test interlnx=24;
run;
Data hw4gold1; set hw4;
if x>610 then delete;
run;
proc reg data=hw4gold1;
model trport = lnx k a;
run;
Data hw4gold2; set hw4;
if x<=610 then delete;
run;
proc reg data=hw4gold2;
model trport = lnx k a;
run;
proc reg data=hw4gold1;
model trportlnx = interlnx klnx alnx;
run;
proc reg data=hw4gold2;
model trportlnx = interlnx klnx alnx;
run;
proc reg data=hw4gold1;
model trportlnx2 = interlnx2 interlnx klnx2 alnx2 /noint;
run;
proc reg data=hw4gold2;
model trportlnx2 = interlnx2 interlnx klnx2 alnx2 /noint;
run;
Data whitehetero; set white;
ehatsqr = ehat1*ehat1;
```

```
a2 = a*a;
k2 = k*k;
lnxa = lnx*a;
lnxk = lnx*k;
ak = a*k;
run;
PROC Reg data=whitehetero;
model ehatsqr = lnx k a lnx2 k2 a2 lnxa lnxk ak;
test lnx=k=a=lnx2=k2=a2=lnxa=lnxk=ak=0;
run;
PROC Autoreg data=hw4;
model trport = lnx k a;
hetero lnx / link=linear test=lm;
run;
PROC Autoreg data=hw4;
model trport = lnx k a / method=ml;
hetero lnx / link=linear test=lm;
run;

PROC Autoreg data=hw4;
model trport = lnx k a;
hetero lnx k a / link=linear test=lm;
run;
PROC Autoreg data=hw4;
model trport = lnx k a / method=ml;
hetero lnx k a/ link=linear test=lm;
run;
```

MAXITER= 1000

ADD THIS, THEN ML FUNCTION WILL CONVERGE.