AAEC 6311

LAB #3

Objectives:

1) Learn to test and correct for heteroskedasticity using SAS

The data used comes from a study of efficiency in production of U.S. airline services (Greene, 2007). Thus, we will fit a cost function for the airline industry:

$$\ln(C_{it}) = \beta_0 + \beta_1 \ln(Q_{it}) + \beta_2 (\ln(Q_{it}))^2 + \beta_3 \ln(PF_{it}) + \beta_4 (LF_{it}) + \varepsilon_{it}, \quad (1)$$

where:

i = Airline,
t = Year,
Q = Output, in revenue passenger miles, index number,
C = Total cost, in $1000,
PF = Fuel price
LF = Load factor, the average capacity utilization of the fleet.

The data comes from 90 observations on 6 firms for 15 Years, 1970-1984.

Part 1. Basic Operations Using SAS

1.1. Import and manipulate the data

1.2. Calculate basic summary statistics and report results:

2. The SAS System

The MEANS Procedure

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
I	**I**	90	3.5000000	1.7173929	1.0000000	6.0000000
T	**T**	90	8.0000000	4.3446984	1.0000000	15.0000000
C	**C**	90	1122523.83	1192074.70	68978.00	4748320.00
Q	**Q**	90	0.5449946	0.5335865	0.0376820	1.9364600
PF	**PF**	90	471683.01	329502.91	103795.00	1015610.00
LF	**LF**	90	0.5604602	0.0527934	0.4320660	0.6762870
lnC		90	13.3656093	1.1319710	11.1415429	15.3733014
lnQ		90	-1.1743092	1.1506057	-3.2785728	0.6608616
lnQ2		90	2.6881856	3.0723943	0	10.7490393
lnPF		90	12.7703594	0.8123749	11.5501731	13.8310000

Part 2. Using proc reg to test for heteroskedasticity and to obtain Heteroskedasticity- Consistent Standard Errors (robust)

2.1. Estimate the parameters of Equation (1) using LS, report regression results:

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: lnC

Number of Observations Read	90
Number of Observations Used	90

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	112.76597	28.19149	1879.55	<.0001
Error	85	1.27492	0.01500		
Corrected Total	89	114.04089			

Root MSE	0.12247	**R-Square**	0.9888
Dependent Mean	13.36561	**Adj R-Sq**	0.9883
Coeff Var	0.91631		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	**1**	9.42058	0.23035	40.90	<.0001
lnQ		**1**	0.93543	0.02929	31.94	<.0001
lnQ2		**1**	0.02254	0.01122	2.01	0.0477
lnPF		**1**	0.45767	0.02004	22.84	<.0001
LF	LF	**1**	-1.53744	0.34232	-4.49	<.0001

2.2 Report the results of proc reg White's test for heteroskedasticity. Interpret the test results.

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: lnC

Heteroscedasticity Consistent Covariance of Estimates

Variable	Label	Intercept	lnQ	lnQ2	lnPF	LF
Intercept	**Intercept**	0.0456617874	0.0014195212	0.000187784	-0.003098434	-0.009436536
lnQ		0.0014195212	0.0007525602	0.0002705379	-0.000023014	-0.002189593
lnQ2		0.000187784	0.0002705379	0.0001095831	-0.000010734	-0.000228033
lnPF		-0.003098434	-0.000023014	-0.000010734	0.0004087039	-0.00373313
LF	**LF**	-0.009436536	-0.002189593	-0.000228033	-0.00373313	0.0988958501

Heteroscedasticity Consistent Covariance of Estimates

Variable	Label	Intercept	lnQ	lnQ2	lnPF	LF

HCC Approximation Method: HC3

Test of First and Second Moment Specification

DF	Chi-Square	Pr > ChiSq
13	30.81	0.0036

In general, why is heteroskedasticity a problem?

Heteroskedasticity means the variance throughout our sample data is not the same. This also means the OLS model is no longer BLUE and that the error variance is biased. We will have incorrect standard errors, invalid t-statistics and F-statistics, and the LM test is no longer valid.

2.4. Compare the standards errors obtained using the 4 versions of White's (Heteroskedasticity-Consistent Covariance Matrix) HCCM and those obtained using the LS Covariance Matrix.

		LS	HCCMETHOD=0	HCCMETHOD=1	HCCMETHOD=2	HCCMETHOD=3
Parameter	Estimate	Standard Error	Standard Error	Standard Error	Standard Error	Standard Error
β_0	9.42058	0.23035	0.20248	0.20835	0.20799	0.21369
β_1	0.93543	0.02929	0.02608	0.02683	0.02674	0.02743
β_2	0.02254	0.01122	0.00994	0.01023	0.01020	0.01047
β_3	0.45767	0.02004	0.01920	0.01976	0.01970	0.02022
β_4	-1.53744	0.34232	0.29634	0.30493	0.30524	0.31448

1) Are the parameter estimates the same obtained using HCCM's different to those obtained using LS? (just check values, you do not need to report them)

   Yes, the parameter estimates obtained from each of the HCCM are the same as the ones obtained using LS.

2) Are the standard errors obtained using HCCM's different to those obtained using LS (from table above)?

   Yes, the standard errors from each of the HCCMETHODs differ from the standard errors obtained using LS.

3) What is the difference between the different versions of the HCCMETHOD? (search online)

- $HC_0 = (X'X)^{-1}(X'\text{diag}(e_i^2)X)(X'X)^{-1}$
- $HC_1 = \frac{n}{n-p} HC_0$
- $HC_2 = (X'X)^{-1}X'\text{diag}(\frac{e_i^2}{1-h_{ii}})X(X'X)^{-1}$ where $h_{ii} = x_i(X'X)^{-1}x_i'$
- $HC_3 = (X'X)^{-1}X'\text{diag}(\frac{e_i^2}{(1-h_{ii})^2})X(X'X)^{-1}$

MacKinnon and White (1985) introduced three alternative heteroscedasticity-consistent covariance matrix estimators that are all asymptotically equivalent to the estimator HC_0 but that typically have better small sample behavior.

2.5. Using proc reg estimate standard errors using bootstrapping

		LS	Boot. Samples=200	Boot. Samples=400	Boot. Samples=500	Boot. Samples=1000
Parameter	Estimate	Standard Error	Standard Error	Standard Error	Standard Error	Standard Error
β_0	9.42058	0.23035	0.2315789	0.2204868	0.2191908	0.2139064
β_1	0.93543	0.02929	0.0282638	0.0272728	0.0269626	0.0276947
β_2	0.02254	0.01122	0.0105374	0.0103506	0.0102192	0.0105177
β_3	0.45767	0.02004	0.0199487	0.0196119	0.0197665	0.0195477
β_4	-1.53744	0.34232	0.2963815	0.3000202	0.3003702	0.306504

Interpret the results above: a) Does it seems necessary to use more than 400 bootstrapping samples? How do the results above compare with the standard errors obtained using HCCM's?

Yes, it seems necessary to use more than 400 bootstrapping samples, since the standard errors still seem to jump around up to 1000 samples. For the most part, the standard errors from the bootstrapping method are similar to the standard errors using each of the 4 versions of White's HCCM models.

Part 3. Using proc model to test for heteroskedasticity and to obtain Heteroskedasticity- Consistent Standard Errors (robust)

3.1. Use **proc model** to estimate the parameters of Equation (1) using LS, report regression results:

The SAS System

The MODEL Procedure

Nonlinear OLS Summary of Residual Errors

Equation	DF Model	DF Error	SSE	MSE	Root MSE	R-Square	Adj R-Sq
lnC	5	85	1.2749	0.0150	0.1225	0.9888	0.9883

Nonlinear OLS Parameter Estimates

Parameter	Estimate	Approx Std Err	t Value	Approx Pr > |t|
bo	9.42058	0.2025	46.53	<.0001
b1	0.935427	0.0261	35.87	<.0001
b2	0.022543	0.00994	2.27	0.0259
b3	0.457668	0.0192	23.83	<.0001
b4	-1.53744	0.2963	-5.19	<.0001

3.2 Report the results of proc model White's test for heteroskedasticity. Interpret the test results.

Heteroscedasticity Test

Equation	Test	Statistic	DF	Pr > ChiSq	Variables
lnC	White's Test	39.45	13	0.0002	Cross of all vars

3.3 Report the results of proc model Breusch-Pagan for heteroskedasticity. Interpret the results. The specific form that we will assume for this test is:

$$\sigma_{it}^2 = \exp\left(\gamma_1 + \gamma_2 LF_{it}\right)$$

The SAS System

The MODEL Procedure

Nonlinear OLS Summary of Residual Errors

Equation	DF Model	DF Error	SSE	MSE	Root MSE	R-Square	Adj R-Sq
lnC	5	85	1.2749	0.0150	0.1225	0.9888	0.9883

Nonlinear OLS Parameter Estimates

Parameter	Estimate	Approx Std Err	t Value	Approx Pr > |t|
bo	9.42058	0.2304	40.90	<.0001
b1	0.935427	0.0293	31.94	<.0001
b2	0.022543	0.0112	2.01	0.0477
b3	0.457668	0.0200	22.84	<.0001
b4	-1.53744	0.3423	-4.49	<.0001

Number of Observations		Statistics for System	
Used	90	**Objective**	0.0142
Missing	0	**Objective*N**	1.2749

Heteroscedasticity Test

Equation	Test	Statistic	DF	Pr > ChiSq	Variables
lnC	White's Test	39.45	13	0.0002	Cross of all vars
	Breusch-Pagan	2.78	1	0.0955	1, LF

3.4. Compare the standards errors obtained using the 4 versions of White's (Heteroskedasticity-Consistent Covariance Matrix) HCCM from **proc model** and those obtained using the LS Covariance Matrix.

Parameter	Estimate	LS Standard Error	HCCME=**0** Standard Error	HCCME=**1** Standard Error	HCCME=**2** Standard Error	HCCME=**3** Standard Error
β_0	9.42058	0.2304	0.2025	0.2083	0.2080	0.2137
β_1	0.935427	0.0293	0.0261	0.0268	0.0267	0.0274
β_2	0.022543	0.0112	0.00994	0.0102	0.0102	0.0105
β_3	0.457668	0.0200	0.0192	0.0198	0.0197	0.0202
β_4	-1.53744	0.3423	0.2963	0.3049	0.3052	0.3145

1) Are the parameter estimates the same? (just check values, you do not need to report them)

Yes, the parameter estimates for the 4 versions of White's HCCM models are the same for every variable.

2) Are the standard errors different (from table above)?

Yes, the standard errors in every case differs from the White's test using just OLS.

3) How do the standard errors obtained using the HCCME option in proc model compare to those obtained using the HCCMETHOD option in proc reg?

   The standard errors using the HCCME option in proc model and proc reg remains the same for each of the HCCME.

4) What is the difference between the different versions of the HCCME option in proc model? (search online)

   - For HCCME=N0, this is just the simple OLS estimator $(x'x)$
   - For HCCME=0, this is calculated as $\frac{1}{M}\sum_{i=0}^{M}\hat{\epsilon}_i^2 x_i x_i'$
   - For HCCME=1, this is calculated as $\frac{1}{M}\sum_{i=0}^{M}(\frac{M}{M-K}\hat{\epsilon}_i^2 x_i x_i')$
   - For HCCME=2, this is calculated as $\frac{1}{M}\sum_{i=0}^{M}\frac{\hat{\epsilon}_i^2}{1-\hat{h}_i}x_i x_i'$
   - For HCCME=3, this is calculated as $\frac{1}{n}\sum_{i=0}^{M}\frac{\hat{\epsilon}_i^2}{(1-\hat{h}_i)^2}x_i x_i'$

Part 4. Estimate White's and Breusch-Pagan Heteroskedasticity tests following the procedures outlined in class which require the estimation of additional regressions using the residuals of the LS procedure to estimate equation (1).

White's test

4.1. Report the regression results of the auxiliary regression needed to obtain White's test

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: ehat2

Number of Observations Read	90
Number of Observations Used	90

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	0.01588	0.00144	5.02	<.0001
Error	78	0.02241	0.00028729		
Corrected Total	89	0.03829			

Root MSE	0.01695	**R-Square**	0.4147
Dependent Mean	0.01417	**Adj R-Sq**	0.3322
Coeff Var	119.65256		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	**1**	-1.11387	0.73633	-1.51	0.1344
lnQ		**1**	0.02351	0.08376	0.28	0.7797
lnQ2		**1**	0.05919	0.02972	1.99	0.0499
lnPF		**1**	0.13699	0.12301	1.11	0.2689
LF	LF	**1**	0.66272	1.05587	0.63	0.5321
lnPF2		**1**	-0.00741	0.00565	-1.31	0.1938
LF2		**1**	-1.68789	1.03250	-1.63	0.1061
lnQlnPF		**1**	-0.00272	0.00714	-0.38	0.7049
lnQLF		**1**	-0.03375	0.09886	-0.34	0.7337
lnQ2lnPF		**1**	-0.00348	0.00319	-1.09	0.2778
lnQ2LF		**1**	-0.05198	0.04186	-1.24	0.2180
lnPFLF		**1**	0.10487	0.09901	1.06	0.2928

4.2. Calculate the nR^2 statistic and compare it with that obtained using proc model.

nR^2=90*0.4147=37.323, which is less than what we obtained using proc model in 3.1 (which was 39.45), but similar.

Breusch-Pagan

4.3. Report the regression results of the auxiliary regression needed to estimate the Breusch-Pagan test.

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: ehat3

Number of Observations Read	90
Number of Observations Used	90

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	5.89179	5.89179	2.80	0.0976
Error	88	184.90190	2.10116		
Corrected Total	89	190.79369			

Root MSE	1.44954	**R-Square**	0.0309
Dependent Mean	1.00000	**Adj R-Sq**	0.0199
Coeff Var	144.95344		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	**1**	-1.73145	1.63831	-1.06	0.2935
LF	LF	**1**	4.87359	2.91041	1.67	0.0976

4.4. Calculate the LM statistic and compare it with that obtained using proc model.

The LM statistic would be ½(5.89179)= 2.945895, which is close to the statistic we got in 3.1 (which was 2.78).

Extra credit

What is the difference between White's test in **proc reg** (spec) and White's t test in **proc model**?

White's test in the MODEL procedure is different than White's test in the REG procedure requested by the SPEC option. The SPEC option produces the test from Theorem 2 on page 823 of White (1980). The WHITE option, on the other hand, produces the statistic discussed in Greene (1993).