CHAPTER
9
HETEROSCEDASTICITY: WHAT HAPPENS IF THE ERROR VARIANCE IS NONCONSTANT?

QUESTIONS

9.1. Heteroscedasticity means that the variance of the error term in a regression model does not remain constant between observations.

(***a***) The OLS estimators are still unbiased but they are no longer efficient.

(***b***) and (***c***) Since the estimated standard errors of OLS estimators may be biased, the resulting *t* ratios are likely to be biased too. As a result, the usual confidence intervals, hypothesis testing procedure, etc. are likely to be of questionable value.

9.2. (***a***) *False*. The OLS estimators are still unbiased; only they are no longer efficient.

(***b***) *True*. Since the estimated standard errors are likely to be biased, the *t* ratios will be biased too.

(***c***) *False*. Sometimes OLS overestimates the variances of OLS estimators and sometimes it underestimates them.

(***d***) *Uncertain*. It may or may not. Sometimes a systematic pattern in the residuals may reflect specification bias, such as omission of a relevant variable, or wrong functional form, etc.

(***e***) *True*. Since the true heteroscedastic variances are not directly observable, one cannot test for heteroscedasticity directly without making some assumptions.

9.3. (***a***) Yes, because of the diversity of firms included in the Fortune 500 list.

(***b***) Probably.

(***c***) Probably not. In time series data, it is often not easy to isolate the effects of autocorrelation and heteroscedasticity.

(***d***) Yes, because of vast differences in per capita income data of developed and developing countries.

(*e*) Yes. Although the U.S. and Canadian inflation rates are similar, the Latin American countries exhibit wide swings in the inflation rate.

9.4. By giving unequal weights, WLS discounts extreme observations. The estimators thus obtained are BLUE. Note that WLS is a specific application of GLS, the method of generalized least squares.

9.5. **(*a*)** This is a visual method, which is often a good starting point to find out if one or more assumptions of the classical linear regression model (CLRM) are fulfilled.

(*b*) and **(*c*)** These two tests formalize the graphical method by making suitable assumptions(s) about the explanatory variable(s) that might be the cause of heteroscedasticity.

PROBLEMS

9.6 Let $Y_i = B_1 + B_2 X_i + u_i$. Now divide this equation through by X_i^2 to obtain:

$$\frac{Y_i}{X_i^2} = B_1 \frac{1}{X_i^2} + B_2 \frac{1}{X_i} + v_i, \text{ where } v_i = \frac{u_i}{X_i^2}$$

The error term v_i is homoscedastic. Use the regression-through-the-origin procedure to estimate the parameters of the transformed model.

9.7. **(*a*)** Perhaps heteroscedasticity is present in the data.

(*b*) $\text{var}(u_i) = \sigma^2(\text{GNP}_i^2)$.

(*c*) The coefficients of the original and transformed models are about the same, although the standard errors of the coefficients in the transformed model seem to be somewhat lower, perhaps suggesting that the authors have succeeded in reducing the severity of heteroscedasticity.

(*d*) No. In the transformed model, the intercept in fact represents the slope coefficient of GNP.

(*e*) The two R^2s cannot be compared directly because the dependent variables in the two models are different.

9.8. (***a***) He is assuming that $\text{var}(u_i) = \sigma^2 X_i$, that is, the error variance is proportional to the distance from the central business district.

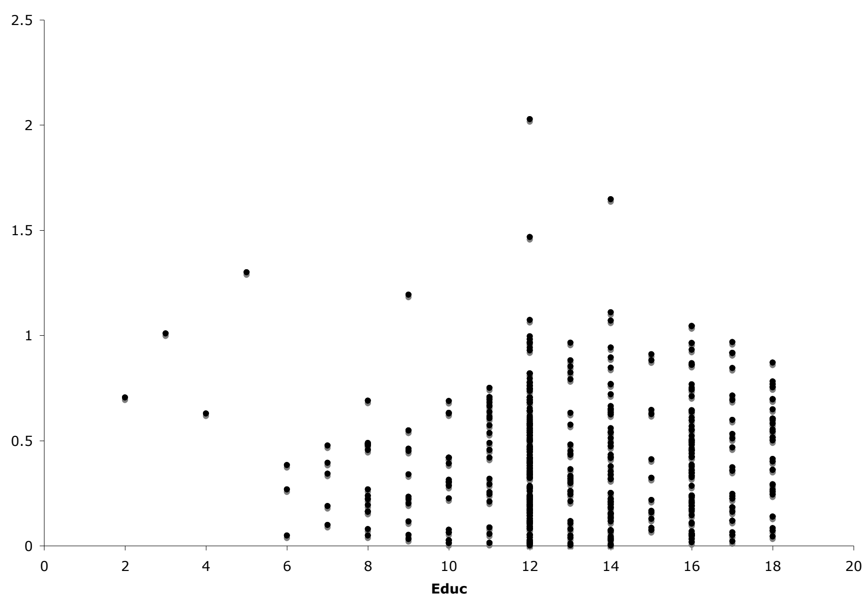
(***b***) Although the values of the slope coefficient in the original and transformed models are about equal, the standard error in the transformed model is lower (i.e., the *t* ratio is higher). This might suggest that the author has probably succeeded in reducing heteroscedasticity.

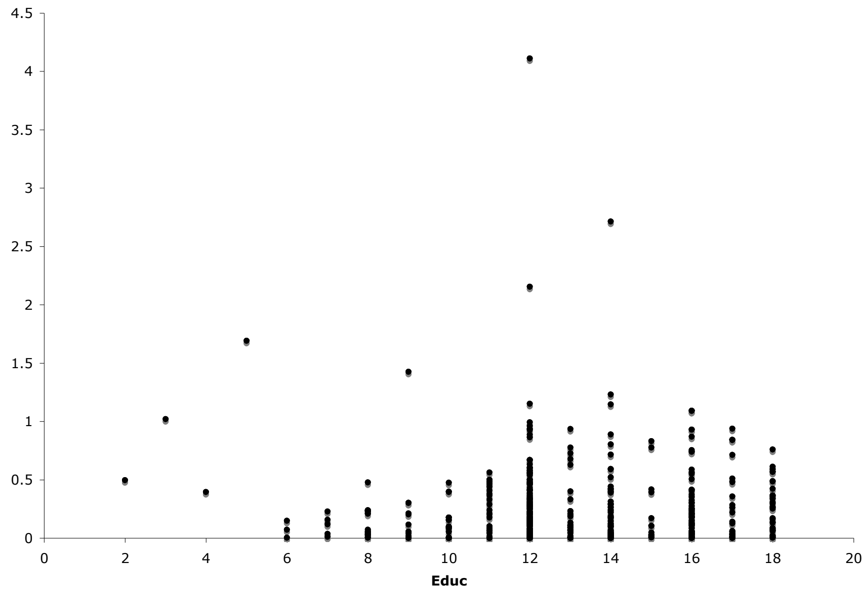
(***c***) The original model is a log-lin model. The slope coefficient of about -0.24 suggests that as the distance traveled from the central business district increases by a mile, the average population density decreases by about 24%. The results make economic sense because the greater the distance one has to travel to get to work, the lesser will be the density of population of that place.

9.9 (***a***) Based on the data in Table 9-2, using ONLY the log of education as the independent variable, the results are as follows:

$$\widehat{\ln Y_i} = 0.0716 + 0.787 \ \ln Educ_i$$
$$\text{se} = (0.2344) \quad (0.0918)$$
$$t = (0.31) \quad (8.57) \qquad\qquad r^2 = 0.124$$

(***b***) The plots indicate the potential of some heteroscedasticity:





(***c***) *Park Test*: $\widehat{\ln e_i^2}$ = -3.080 + 0.1762 ln($\ln Educ_i$)

t = (-2.91) (0.43) $r^2 = 0.00$

Note: The original regression is double-logarithmic. Therefore, in the Park test we are using the natural log of the squared residuals and the natural log of ln*Educ*.

Since the slope coefficient in this regression is not statistically significant at the 5% level, the Park test does not suggest the presence of heteroscedasticity.

Glejser Test: $|e_i|$ = 0.3593 + 0.0151 $\ln Educ_i$

t = (2.64) (0.28) $r^2 = 0.000$

This particular form of the Glejser test suggests that there is no heteroscedasticity.

(***d***) In the present case, the question is academic.

(***e***) Perhaps the log-linear model.

(***f***) No, because the dependent variables in the two models are not the same.

9.10. (***a***) $\widehat{\text{Wage}}_i$ = 8.6406 + 0.0263 Exper $_i$

t = (21.44) (1.43) $r^2 = 0.04$

$\ln \widehat{\text{Wage}}_i$ = 1.8244 + 0.0951 ln Exper $_i$

t = (25.47) (3.65) $r^2 = 0.025$

(***b***) In the linear model there seems to be some evidence of heteroscedasticity. In the log-linear model such evidence is not clear.

(***c***) *Linear Model*:

(1) *Park Test*: $\widehat{\ln e_i^2}$ = -7.8314 + 2.4958 ln Exper$_i$

t = (6.98) (-0.96) r^2 = 0.002

Since the estimated t value is not significant, the Park test does suggest heteroscedasticity.

(2) *Glejser test*: $|e_i|$ = 3.723 + 0.0059 Exper$_i$

t = (13.90) (0.48) r^2 = 0.000

Again, there is no indication of heteroscedasticity, since the estimated t value of Experience is not statistically significant. If you repeat the Park and Glejser tests for the log-linear model, you will find that the regression results are not significant.

(***d***) Since there was no evidence of heteroscedasticity, this is left to the reader.

9.11. (***a***) Let Y = GDP growth rate (%), and $X = \dfrac{\text{Investment}}{\text{GDP}}$ (%).

You can regress Y on X. You can also regress ln Y on ln X, provided the Y values are positive. To make all the Y values positive, add a constant in such a way that the largest negative value becomes positive.

(***b***) Yes, there is evidence of heteroscedasticity. This should not be surprising because the countries in the sample have positive as well as negative real interest rates.

(***c***) If it is assumed that the error variance is proportional to the value of X, use the square root transformation. If it is assumed that the error variance is proportional to the square of X, divide the equation by X on both sides.

(***d***) Add two dummy variables to the model to distinguish the three categories of interest rate experiences. If the original model (without the dummies) was mis-specified, and if the residuals in the new model (i.e., with the dummies added) do not exhibit any systematic pattern, the "heteroscedasticity" observed in the original model can then be attributed to the mis-specification bias.

9.12. Let Y = median salary and X = age (Assume $X = 72$ for the last group).

(***a***) $\hat{Y} = 6{,}419.8182 + 127.8182\,X$

$t =$ (3.6408) (3.5946) $r^2 = 0.5894$

(***b***) $\dfrac{Y}{\sqrt{X}} = 5{,}133.8548\left(\dfrac{1}{\sqrt{X}}\right) + 155.1791\sqrt{X}$

$t =$ (3.6702) (4.8764) $r^2 = 0.9608$

Note: This is a regression without an intercept. The r^2 shown is based on the raw r^2 formula. The original r^2 in *EViews* is negative, a common occurrence when the intercept is suppressed.

(***c***) $\dfrac{Y}{X} = 4{,}216.9105\left(\dfrac{1}{X}\right) + 177.4836$

$t =$ (3.8596) (6.2138) $r^2 = 0.6234$

(***d***) It seems that transformations (*b*) and (*c*) have reduced the standard errors in relation to the coefficients, probably reducing the heteroscedasticity problem. Plot the residuals from regressions (*b*) and (*c*) and see if they exhibit any systematic patterns. If they do, use the Park or Glejser test to further confirm if there is evidence of heteroscedasticity in the data.

9.13. The Spearman's rank correlation coefficient is 0.4407. Substituting this value in the given formula, the *t* value is 1.9636. For 16 d.f., the 5% one-tailed critical *t* value is 1.746. Therefore, the observed *t* value is significant at this level, suggesting perhaps that there is evidence of heteroscedasticity in the data.

9.14. (***a***) $\hat{Y}_i = 1{,}993.7258 + 0.2328\,X_i$

$t =$ (2.1309) (2.3340) $r^2 = 0.4376$

(***b***) $\dfrac{Y_i}{\sigma_i} = 2{,}417.3347\left(\dfrac{1}{\sigma_i}\right) + 0.1800\left(\dfrac{X_i}{\sigma_i}\right)$

$$t = (2.1131) \qquad (1.4273) \qquad r^2 = 0.6482$$

Note: This r^2 is the one generated by *EViews*. Since the regression does not have an intercept, you may wish to calculate the raw r^2 as an exercise.

In this example, the unweighted regression may be more appropriate based on the statistical significance of the coefficients.

9.15. $$\text{var}(v_i) = \text{var}\left(\frac{u_i}{X_i}\right) = \frac{\text{var}(u_i)}{X_i^2} = \frac{\sigma^2 X_i^2}{X_i^2} = \sigma^2$$

9.16. (***a***) In regression (1) the slope coefficient suggests that if the number of employees increases by 1, the average salary goes up by 0.009 dollars. After multiplying through by *N*, the slope coefficient in model (2) is about the same as in model (1).

(***b***) The author is not only assuming heteroscedasticity, but specifically states that the error variance is proportional to the square of *N*.

(***c***) As noted in (*a*), the two slopes and the two intercepts are about the same.

(***d***) Because the two dependent variables are not the same, the two R^2s cannot be compared directly.

9.17. The derived average and marginal cost functions are as follows:

Average cost function [From Eq. (9.32)]: $\left(\frac{Y_i}{X_i}\right) = 476{,}000\left(\frac{1}{X_i}\right) + 31.348 - (1.083 \times 10^{-6})X_i$
Marginal cost function [from Eq.(9.32)]: $\left(\frac{dY_i}{dX_i}\right) = 31.348 - 2(1.083 \times 10^{-6})X_i$
Average cost function [from Eq.(9.33)]: $\left(\frac{Y_i}{X_i}\right) = 342{,}000\left(\frac{1}{X_i}\right) + 25.57 + (4.34 \times 10^{-6})X_i$
Marginal cost function [from Eq. (9.33]:

$$\left(\frac{dY_i}{dX_i}\right) = 25.57 + 2(4.34 \times 10^{-6})X_i$$

In Model (9.33) the quadratic term in X is not statistically significant, suggesting that the total cost function is linear. This means the average and marginal cost functions derived from (9.33) are in fact:

$$\text{Average cost} = \frac{342{,}000}{X} + 25.57$$

$$\text{Marginal cost} = 25.57$$

Note: If you need to refresh your memory on the concepts of various cost functions, consult any introductory microeconomics textbook.

9.18. (***a***) *A priori,* calorie intake should have a negative effect on infant mortality and population growth should have a positive effect.

(***b***) The *EViews* regression results are as follows:

(Regression output is shown on the following page)

Dependent Variable: IMOR
Sample: 1 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	172.6195	52.45598	3.290749	0.0050
PCGNP	-0.002502	0.001535	-1.629641	0.1240
PEDU	-1.279618	0.316722	-4.040198	0.0011
POPGROWTH	6.379603	7.045706	0.905460	0.3795
CSPC	-0.001363	0.018708	-0.072873	0.9429
R-squared	0.815002	F-statistic		16.52053
Adjusted R-squared	0.765670	Prob(F-statistic)		0.000023

Note: We are showing the F statistic and its p value here.

The population growth and calorie intake variables have the expected signs.

(***c***) Only one of the coefficients in the preceding regression is statistically significant, yet the F value is very significant. This seems to be a classic case of multicollinearity. Dropping the population growth (POPGROWTH) and per capita GNP (PCGNP) variables, the results were as follows:

Dependent Variable: IMOR Sample: 1 20				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	250.0996	28.19164	8.871411	0.0000
PEDU	-1.210991	0.329756	-3.672385	0.0019
CSPC	-0.030999	0.013099	-2.366448	0.0301
R-squared	0.760462	F-statistic		26.98494
Adjusted R-squared	0.732281	Prob(F-statistic)		0.000005

Now both independent variables are statistically significant.

9.19. (***a***) The regression results show that the none of the coefficients in the auxiliary regression are statistically significant.

(***b***) Since not only the coefficients are insignificant but also the product of the R^2 and the sample size will not exceed the critical χ^2 value at 5 d.f., we can conclude there is no evidence of heteroscedasticity.

(***c***) Examine the residuals from the transformed model visually. You can also apply the White procedure to the residuals from the transformed regressions to make sure that they are not heteroscedastic.

9.20. (***a***) To explain the caloric intake, a model using the variables per capita GNP (PCGNP, or X_2), index of literacy (PEDU, or X_3), and population growth (POPGROWTH, or X_4) was developed. X_4 was insignificant and was dropped from the model, and the final *EViews* model was as follows:

Dependent Variable: CSPC Sample: 1 20				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1497.320	313.2984	4.779214	0.0002
PCGNP	0.060858	0.014767	4.121147	0.0007
PEDU	10.61977	3.556400	2.986102	0.0083
R-squared	0.716941			

(***b***) When plotted against the independent variables, the residuals from the preceding regression model showed visible heteroscedastic patterns.

(***c***) Using *EViews*, the following White's heteroscedasticity-corrected regression was obtained:

Dependent Variable: CSPC Sample: 1 20 White Heteroskedasticity-Consistent Standard Errors & Covariance				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1497.320	333.8182	4.485436	0.0003
PCGNP	0.060858	0.009457	6.435356	0.0000
PEDU	10.61977	3.855596	2.754379	0.0135
R-squared	0.716941			

As you can see comparing this regression with the one given in (*a*), the standard errors using the White procedure are different, in one case much lower and in the other a bit higher. That is, this procedure gives more efficient estimates of the parameters while allowing us to retain the original regression estimates.

9.21. Consider Model 1 in Table 7-2. Applying White's heteroscedasticity test (with no cross-product terms), we get the following results from *EViews*:

(Regression output is shown on the following page)

White Heteroskedasticity Test:				
F-statistic	3.214078	Probability		0.016850
Obs*R-squared	11.76858	Probability		0.019158
Test Equation: Dependent Variable: RESID^2 Sample: 1 85				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-42.97513	32.53147	-1.321033	0.1903
INCOME	0.002603	0.004075	0.638917	0.5247
INCOME^2	-1.72E-07	2.13E-07	-0.804801	0.4233
ACCESS	2.807384	1.124675	2.496174	0.0146
ACCESS^2	-0.023232	0.009280	-2.503543	0.0143
R-squared	0.138454			

Note: RESID^2 means residuals squared, and so on. The White n^2 test statistic is also shown (Obs*R-squared), and it is significant (its p value is 0.0191). Incidentally, even if we introduce the cross-product terms, there is evidence of heteroscedasticity. When running the initial regression, do not forget to save your residuals in a new series so that you can apply the White test or other tests: The RESID series in each *EViews* work file is used as a depository of the residuals from each regression you run, and each new regression overwrites the residuals of the previous one.

These results suggest that we have a heteroscedasticity problem. One can use a variety of transformations to resolve it. You are urged to plot the squared residuals of the chosen model on each of the explanatory variables and / or on the estimated values of the dependent variable to see which variable might be used to transform the data to eliminate heteroscedasticity. We will give here the results of White's heteroscedasticity-corrected standard errors for Model 1 of Table 7.2, which are as follows:

Dependent Variable: LE
Sample: 1 85
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	39.43802	1.823039	21.63313	0.0000
INCOME	0.000542	9.52E-05	5.695746	0.0000
ACCESS	0.283303	0.026132	10.84117	0.0000
R-squared	0.774146			

A comparison with the results given in Table 7-2 will show that apparently the original model overestimated the standard errors, for the estimated t values are lower in that table as compared with the t values shown in the preceding regression.

You can proceed similarly with the remaining two models in Table 7-2.

9.22. For the Experience variable:

$$|e_i| = 3.165 + 0.0108 Exper$$
$$t = (12.77) \quad (0.95) \qquad r^2 = 0.002$$

$$|e_i| = 3.0989 + 0.0655\sqrt{Exper}$$
$$t = (7.79) \ (0.70) \qquad r^2 = 0.001$$

$$|e_i| = 3.2480 + 0.9528\frac{1}{Exper}$$
$$t = (19.19) \ (1.15) \qquad r^2 = 0.003$$

For the Wage variable:

$$|e_i| = 0.3390 + 0.3314 Wage$$
$$t = (1.43) \ (14.63) \qquad r^2 = 0.291$$

$$|e_i| = -1.5942 + 1.6988\sqrt{Wage}$$
$$t = (-3.29) \ (10.59) \qquad r^2 = 0.177$$

$$|e_i| = 3.7773 - 2.894\frac{1}{Wage}$$
$$t = (13.38) \ (-1.69) \qquad r^2 = 0.005$$

9.23. The Breusch-Pagan test assesses whether the error variance is a function of one or more of the independent variables. The ratio of each squared residual over the ML estimator of the error variance is saved; this new column is now regressed against the (m – 1) chosen independent variables. The (1/2)ESS value is approximately a Chi-squared variable with (m – 1) degrees of freedom. Verification of heteroscedasticity in (9.33) is:

```
The regression equation is
p = 0.171 + 0.0692 Educ + 0.0106 Exper - 0.221 Sex - 0.209 Marstat
    + 0.093 Region - 0.268 Union

Predictor       Coef    SE Coef       T       P
Constant      0.1706     0.5120    0.33   0.739
Educ         0.06917    0.03289    2.10   0.036
Exper       0.010558   0.007293    1.45   0.148
Sex          -0.2209     0.1631   -1.35   0.176
Marstat      -0.2091     0.1753   -1.19   0.233
Region        0.0930     0.1783    0.52   0.602
Union        -0.2684     0.2115   -1.27   0.205

S = 1.82416   R-Sq = 1.7%   R-Sq(adj) = 0.5%

Analysis of Variance

Source            DF        SS      MS      F       P
Regression         6    28.893   4.816   1.45   0.195
```

Residual Error	516	1717.021	3.328
Total	522	1745.914	

The explained sum of squares is 28.893, so (1/2)(28.893) = 14.4465. With 6 degrees of freedom, the 5% critical value for the Chi-square is 12.5916, and the 1% critical value is 16.8119. So, at the 5% level there is slight evidence of some heteroscedasticity, but at the 1% level we cannot reject the null hypothesis of homoscedasticity.

9.24. The estimation is left to the reader.

9.25. In model (9.33), there are 3 dummy variables. For the SEX variable, the coefficient suggests that women make, on average, about 24.4% less than men with similar characteristics. If a person is married, he or she is likely to make about 6.9% more than similar unmarried people, and if an employee belongs to a union, he or she makes about 18.36% more than similar non-union workers.

9.26 ***(a)*** The linear-in-variables results from *Eviews* are:

Dependent Variable: DOMESTIC
Method: Least Squares
Sample: 1 14
Included observations: 14

	Coefficient	Std. Error	t-Statistic	Prob.
C	10785.52	25182.56	0.428293	0.6767
R_D	3.257962	1.283377	2.538586	0.0275
PROFITS	1.304895	0.222216	5.872188	0.0001

R-squared	0.799630	Mean dependent var	157751.1
Adjusted R-squared	0.763200	S.D. dependent var	91645.36
S.E. of regression	44596.59	Akaike info criterion	24.43611
Sum squared resid	2.19E+10	Schwarz criterion	24.57305
Log likelihood	-168.0528	Hannan-Quinn criter.	24.42344
F-statistic	21.94928	Durbin-Watson stat	2.029377
Prob(F-statistic)	0.000145		

(b) Using White's heteroscedasticity-corrected procedure, the new results are:

Dependent Variable: DOMESTIC
Method: Least Squares
Sample: 1 14
Included observations: 14
White Heteroskedasticity-Consistent Standard Errors & Covariance

	Coefficient	Std. Error	t-Statistic	Prob.
C	10785.52	26195.63	0.411730	0.6884
R_D	3.257962	0.988018	3.297471	0.0071
PROFITS	1.304895	0.252824	5.161281	0.0003

R-squared	0.799630	Mean dependent var	157751.1
Adjusted R-squared	0.763200	S.D. dependent var	91645.36
S.E. of regression	44596.59	Akaike info criterion	24.43611
Sum squared resid	2.19E+10	Schwarz criterion	24.57305
Log likelihood	-168.0528	Hannan-Quinn criter.	24.42344
F-statistic	21.94928	Durbin-Watson stat	2.029377
Prob(F-statistic)	0.000145		

(c) Note that the coefficients have remained unchanged between the two outputs. With respect to standard errors and statistical significance, however, there is a slight difference. The White's corrected results indicate an even stronger level of significance for the R&D variable (its p-value went from 0.0275 to 0.0071). Oddly enough, the Profits p-value went from 0.0001 to 0.0003, slightly decreasing in significance, although it remains extremely useful nonetheless.

9.27. *(a)* Regression results from *EViews* are as follows:

Dependent Variable: SALARY
Sample: 1 447
Included observations: 447

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	998.7095	623.6954	1.601277	0.1100
TENURE	31.67279	9.465097	3.346272	0.0009
AGE	5.492393	11.46086	0.479230	0.6320
SALES	0.014287	0.006614	2.160040	0.0313
PROFITS	0.141302	0.068845	2.052471	0.0407
ASSETS	0.007630	0.001326	5.754849	0.0000

R-squared	0.248829	Mean dependent var	2027.517
Adjusted R-squared	0.240312	S.D. dependent var	1722.566
S.E. of regression	1501.390	Akaike info criterion	17.47950

Sum squared resid	9.94E+08	Schwarz criterion	17.53457
Log likelihood	-3900.669	F-statistic	29.21660
Durbin-Watson stat	2.014806	Prob(F-statistic)	0.000000

The results for the White Heteroskedasticity test are:

White Heteroskedasticity Test:

F-statistic	2.089313	Probability	0.024109
Obs*R-squared	20.44073	Probability	0.025349

With a *p value* of 0.0253, there is apparent heteroscedasticity in the data. It is left as an exercise to the reader to construct the Breusch-Pagan statistic, which also indicates heteroscedasticity in this dataset.

(b) Results for the log-lin model and White's heteroscedasticity test are as follows:

Dependent Variable: LN_SAL
Sample: 1 447
Included observations: 447

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.753659	0.236823	28.51778	0.0000
TENURE	0.008251	0.003594	2.295836	0.0222
AGE	0.007228	0.004352	1.661010	0.0974
SALES	6.09E-06	2.51E-06	2.425693	0.0157
PROFITS	5.72E-05	2.61E-05	2.186738	0.0293
ASSETS	2.03E-06	5.03E-07	4.035537	0.0001

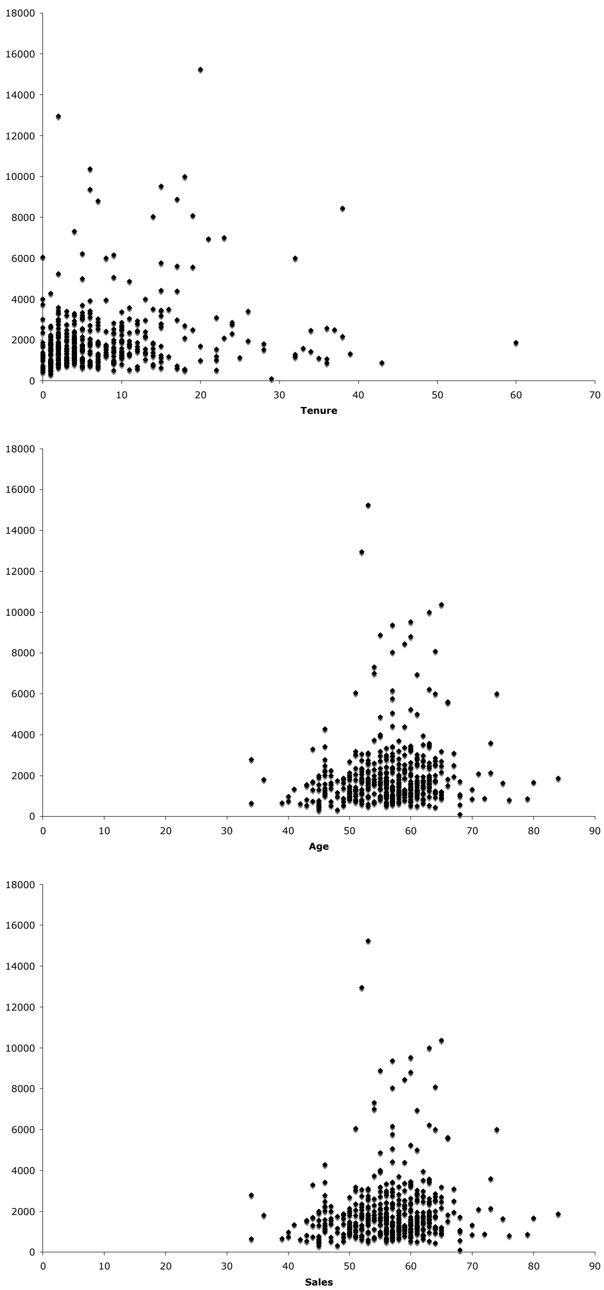
R-squared	0.208984	Mean dependent var	7.391898
Adjusted R-squared	0.200016	S.D. dependent var	0.637388
S.E. of regression	0.570091	Akaike info criterion	1.727292
Sum squared resid	143.3268	Schwarz criterion	1.782359
Log likelihood	-380.0497	F-statistic	23.30217
Durbin-Watson stat	1.920217	Prob(F-statistic)	0.000000

White Heteroskedasticity Test:

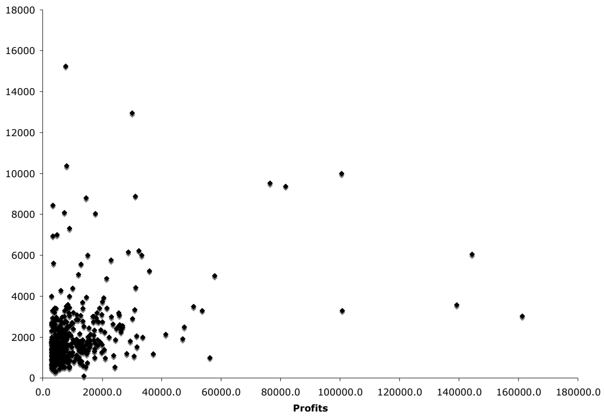
F-statistic	2.581930	Probability	0.004784
Obs*R-squared	24.99078	Probability	0.005363

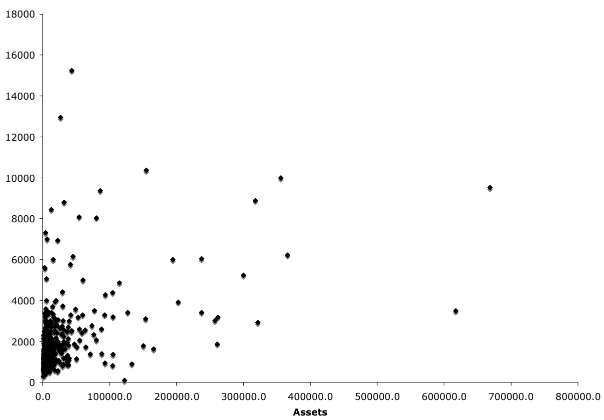
Apparently there is still some heteroscedasticity in the data.

(c)



18000
16000
14000
12000
10000
8000
6000
4000
2000
0
0
10
20
30
40
50
60
70
Tenure
18000
16000
14000
12000
10000
8000
6000
4000
2000
0
0
10
20
30
40
50
60
70
80
90
Age
18000
16000
14000
12000
10000
8000
6000
4000
2000
0
0
10
20
30
40
50
60
70
80
90
Sales





Based on these scattergrams, there are several variables that might be adding to the heteroscedasticity. It is left to the reader to try several models to see which helps decrease it sufficiently.

9.28. (***a***) The regression results are as follows:

$$
\begin{aligned}
\hat{MPG}_i &= 189.9597 - 1.2716SP_i + 0.3904HP_i - 1.9032WT_i \\
se &= (22.5287) \quad (0.2331) \quad (0.0762) \quad (0.1855) \\
t &= (8.4318) \quad (-5.4551) \quad (5.1207) \quad (-10.2593) \\
& \qquad R^2 = 0.8828
\end{aligned}
$$

As expected, MPG is positively related to HP and negatively related to speed and weight.

(***b***) Since this is a cross-sectional data involving a diversity of cars, a *priori* one would expect heteroscedasticity.

(***c***) Regressing the squared residuals obtained from the model shown in (*a*) on the three regressors, their squared terms, and their cross-product terms, we obtain an R^2 value of 0.3094. Multiplying this value by the number of observations (=81), we obtain 25.0646, which under the null hypothesis that there is no heteroscedasticity, has the Chi-square distribution with 9 d.f. (3 regressors, 3 squared regressors, and 3 three cross-product terms). The *p value* of obtaining a Chi-square value of as much as 25.0646 or greater (under the null hypothesis) is 0.0029, which is very small. Hence, we must reject the null hypothesis. That is, there is heteroscedasticity.

(d) The results based on White's procedure are as follows:

Dependent Variable: MPG
Method: Least Squares

Sample: 1 81
Included observations: 81
White Heteroscedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	189.9597	33.90605	5.602531	0.0000
SP	-1.271697	0.336039	-3.784375	0.0003
HP	0.390433	0.108781	3.589180	0.0006
WT	-1.903273	0.285077	-6.676352	0.0000

R-squared 0.882864; Durbin-Watson 1.0237

When you compare this result with the OLS results, you will find that the values of the estimated coefficients are the same, but their variances and standard errors are different. As you can see, the standard errors of all the estimated slope coefficients are higher under the White procedure, hence $|t|$ are lower, suggesting that OLS had underestimated the standard errors. This could all be due to heteroscedasticity.

(e) There is no simple formula to determine the exact nature of heteroscedasticity in the present case. Perhaps one could make some simple assumptions and try various transformations. For example, if it is believed that the "culprit" variable is HP, and if we believe that the error variance is proportional to the square of HP, we could divide through by HP and see what happens. Of course, any other regressor is a likely candidate for transformation.