Chapter 11

Heteroskedasticity

11.1 The Nature of Heteroskedasticity

- In Chapter 3 we introduced the linear model

$$y = \beta_1 + \beta_2 x \tag{11.1.1}$$

to explain household expenditure on food (y) as a function of household income (x).

- We begin this section by asking whether a function such as $y = \beta_1 + \beta_2 x$ is better at explaining expenditure on food for low-income households than it is for high-income households.
- Income is less important as an explanatory variable for food expenditure of high-income families. It is harder to guess their food expenditure.

- This type of effect can be captured by a statistical model that exhibits heteroskedasticity.

$$y_t = \beta_1 + \beta_2 x_t + e_t \tag{11.1.2}$$

- We assumed the e_t were uncorrelated random error terms with mean zero and constant variance σ^2. That is,

$$E(e_t) = 0 \qquad \text{var}(e_t) = \sigma^2 \qquad \text{cov}(e_i, e_j) = 0 \tag{11.1.3}$$

- Including the standard errors for b_1 and b_2, the estimated mean function was

$$\hat{y}_t = \underset{(22.139)}{40.768} + \underset{(0.0305)}{0.1283}\ x_t \tag{11.1.4}$$

- A graph of this estimated function, along with all the observed expenditure-income points (y_t, x_t), appears in Figure 11.1.

- Notice that, as income (x_t) grows, the observed data points (y_t, x_t) have a tendency to deviate more and more from the estimated mean function.

- The least squares residuals, defined by

$$\hat{e}_t = y_t - b_1 - b_2 x_t \tag{11.1.5}$$

increase in absolute value as income grows.

[Figure 11.1 here]

- The observable least squares residuals $(\hat{e}_t)$ are proxies for the unobservable errors (e_t) that are given by

$$e_t = y_t - \beta_1 - \beta_2 x_t \tag{11.1.6}$$

- The information in Figure 11.1 suggests that the unobservable errors also increase in absolute value as income (x_t) increases.
- Is this type of behavior consistent with the assumptions of our model?

- The parameter that controls the spread of y_t around the mean function, and measures the uncertainty in the regression model, is the variance σ^2.
- If the scatter of y_t around the mean function increases as x_t increases, then the uncertainty about y_t increases as x_t increases, and we have evidence to suggest that the variance is not constant.
- Thus, we are questioning the constant variance assumption

$$\operatorname{var}(y_t) = \operatorname{var}(e_t) = \sigma^2 \tag{11.1.7}$$

- The most general way to relax this assumption is to add a subscript t to σ^2, recognizing that the variance can be different for different observations. We then have

$$\operatorname{var}(y_t) = \operatorname{var}(e_t) = \sigma_t^2 \tag{11.1.8}$$

- In this case, when the variances for all observations are not the same, we say that *heteroskedasticity* exists. Alternatively, we say the random variable y_t and the random error e_t are *heteroskedastic*.

- Conversely, if (11.1.7) holds we say that *homoskedasticity* exists, and y_t and e_t are *homoskedastic*.
- The heteroskedastic assumption is illustrated in Figure 11.2.

[Figure 11.2 here]

The existence of different variances, or heteroskedasticity, is often encountered when using *cross-sectional data*.

11.2 The Consequences of Heteroskedasticity for the Least Squares Estimator

- If we have a linear regression model with heteroskedasticity and we use the least squares estimator to estimate the unknown coefficients, then:

1. The least squares estimator is still a linear and unbiased estimator, but it is no longer best. It is no longer B.L.U.E.

2. The standard errors usually computed for the least squares estimator are incorrect. Confidence intervals and hypothesis tests that use these standard errors may be misleading.

- Consider the model

$$y_t = \beta_1 + \beta_2 x_t + e_t \qquad (11.2.1)$$

where

$$E(e_t) = 0 \quad \text{var}(e_t) = \sigma_t^2 \quad \text{cov}(e_i, e_j) = 0 \qquad (i \neq j)$$

- In Chapter 4, equation 4.2.1, we wrote the least squares estimator for β_2 as

$$b_2 = \beta_2 + \sum w_t e_t \qquad (11.2.2)$$

where

$$w_t = \frac{x_t - \bar{x}}{\sum (x_t - \bar{x})^2}$$

- The first property that we establish is that of unbiasedness.

$$\begin{aligned} E(b_2) &= E(\beta_2) + E\left(\sum w_t e_t\right) \\ &= \beta_2 + \sum w_t E(e_t) = \beta_2 \end{aligned} \tag{11.2.4}$$

- The next result is that the least squares estimator is no longer best. The way we tackle this question is to derive an alternative estimator which *is* the best linear unbiased estimator. This new estimator is considered in Sections 10.3 and 11.5.
- To show that the usual formulas for the least squares standard errors are incorrect under heteroskedasticity, we return to the derivation of var(b_2) in (4.2.11). From that equation, and using (11.2.2), we have

$$
\begin{aligned}
\operatorname{var}(b_2) &= \operatorname{var}(\beta_2) + \operatorname{var}\left(\sum w_t e_t\right) \\
&= \operatorname{var}\left(\sum w_t e_t\right) \\
&= \sum w_t^2 \operatorname{var}(e_t) + \sum_{i \neq j} \sum w_i w_j \operatorname{cov}(e_i, e_j) \\
&= \sum w_t^2 \sigma_t^2
\end{aligned}
$$

$$
= \frac{\sum \left[(x_t - \bar{x})^2 \sigma_t^2 \right]}{\left[\sum (x_t - \bar{x})^2 \right]^2} \tag{11.2.5}
$$

- Note from the last line in (11.2.5) that

$$
\operatorname{var}(b_2) \neq \frac{\sigma^2}{\sum (x_t - \bar{x})^2} \tag{11.2.6}
$$

- Note that standard computer software for least squares regression will compute the estimated variance for b_2 based on (11.2.6), unless told otherwise.

11.2.1White's Approximate Estimator for the Variance of the Least Squares Estimator

- Halbert White, an econometrician, has suggested an estimator for the variances and covariances of the least squares coefficient estimators when heteroskedasticity exists.
- In the context of the simple regression model, his estimator for var(b_2) is obtained by replacing σ_t^2 by the squares of the least squares residuals $\hat{e}_t^2$, in (11.2.5).
- Large variances are likely to lead to large values of the squared residuals.
- Because the squared residuals are used to approximate the variances, White's estimator is strictly appropriate only in large samples.

- If we apply White's estimator to the food expenditure-income data, we obtain

$$\hat{\text{var}}(b_1) = 561.89 \qquad \hat{\text{var}}(b_2) = 0.0014569$$

- We could write our estimated equation as

$$\begin{array}{llll} \hat{y}_t = 40.768 & + & 0.1283\, x_t & \\ \quad (23.704) & & (0.0382) & \text{(White)} \\ \quad (22.139) & & (0.0305) & \text{(incorrect)} \end{array}$$

- In this case, ignoring heteroskedasticity and using incorrect standard errors tends to overstate the precision of estimation; we tend to get confidence intervals that are narrower than they should be.

- We can construct two corresponding 95% confidence intervals for β_2.

White: $b_2 \pm t_c \text{se}(b_2) = 0.1283 \pm 2.024(0.0382) = [0.051, 0.206]$

Incorrect: $b_2 \pm t_c \text{se}(b_2) = 0.1283 \pm 2.024(0.0305) = [0.067, 0.190]$

11.3 Proportional Heteroskedasticity

- Return to the example where weekly food expenditure (y_t) is related to weekly income (x_t) through the equation

$$y_t = \beta_1 + \beta_2 x_t + e_t \qquad (11.3.1)$$

- We make the following assumptions:

$$E(e_t) = 0 \quad \text{var}(e_t) = \sigma_t^2$$

$$\text{cov}(e_i, e_j) = 0 \;\; (i \neq j)$$

- By itself, the assumption var(e_t) = σ_t^2 is not adequate for developing a better procedure for estimating β_1 and β_2.

- We overcome this problem by making a further assumption about the σ_t^2. Our earlier inspection of the least squares residuals suggested that the error variance increases as income increases. A reasonable model for such a variance relationship is

$$\operatorname{var}(e_t) = \sigma_t^2 = \sigma^2 x_t \tag{11.3.2}$$

- The assumption of heteroskedastic errors in (11.3.2) is a reasonable one for the expenditure model.
- Under heteroskedasticity the least squares estimator is not the best linear unbiased estimator. One way of overcoming this dilemma is *to change or transform our statistical model into one with homoskedastic errors*. Leaving the basic structure of the model intact, it is possible to turn the heteroskedastic error model into a homoskedastic error model. Once this transformation has been carried out, application of least squares to the transformed model gives a best linear unbiased estimator.
- Begin by dividing both sides of the original equation in (11.3.1) by $\sqrt{x_t}$

$$\frac{y_t}{\sqrt{x_t}} = \beta_1 \frac{1}{\sqrt{x_t}} + \beta_2 \frac{x_t}{\sqrt{x_t}} + \frac{e_t}{\sqrt{x_t}} \tag{11.3.3}$$

- Define the *transformed variables*

$$y_t^* = \frac{y_t}{\sqrt{x_t}} \quad x_{t1}^* = \frac{1}{\sqrt{x_t}} \qquad x_{t2}^* = \frac{x_t}{\sqrt{x_t}} \quad e_t^* = \frac{e_t}{\sqrt{x_t}} \tag{11.3.4}$$

- (11.3.3) can be rewritten as

$$y_t^* = \beta_1 x_{t1}^* + \beta_2 x_{t2}^* + e_t^* \tag{11.3.5}$$

- The beauty of this transformed model is that the new transformed error term e_t^* is homoskedastic. The proof of this result is:

$$\text{var}(e_t^*) = \text{var}\left(\frac{e_t}{\sqrt{x_t}}\right) = \frac{1}{x_t}\text{var}(e_t) = \frac{1}{x_t}\sigma^2 x_t = \sigma^2 \tag{11.3.6}$$

- The transformed error term will retain the properties $E(e_t^*) = 0$ and zero correlation between different observations, $\operatorname{cov}(e_i^*, e_j^*) = 0$ for $i \neq j$.
- As a consequence, we can apply least squares to the transformed variables, y_t^*, x_{t1}^* and x_{t2}^* to obtain the best linear unbiased estimator for β_1 and β_2.
- The transformed model is linear in the unknown parameters β_1 and β_2. These are the original parameters that we are interested in estimating.
- The transformed model satisfies the conditions of the Gauss-Markov Theorem, and the least squares estimators defined in terms of the transformed variables are B.L.U.E.
- The estimator obtained in this way is called a generalized least squares estimator.
- One way of viewing the generalized least squares estimator is as a *weighted least squares estimator*. Recall that the least squares estimator is those values of β_1 and β_2 that minimize the sum of squared errors. In this case, we are minimizing the sum of squared transformed errors that are given by

$$\sum_{t=1}^{T} e_t^{*2} = \sum_{t=1}^{T} \frac{e_t^2}{x_t}$$

- The errors are *weighted* by the reciprocal of x_t. When x_t is small, the data contain more information about the regression function and the observations are weighted heavily. When x_t is large, the data contain less information and the observations are weighted lightly. In this way we take advantage of the heteroskedasticity to improve parameter estimation.

Remark: In the transformed model $x_{t1}^* \neq 1$. That is, the variable associated with the intercept parameter is no longer equal to "1". Since least squares software usually automatically inserts a "1" for the intercept, when dealing with transformed variables you will need to learn how to turn this option off. If you use a "weighted" or "generalized" least squares option on your software, the computer will do both the transforming and the estimating. In this case suppressing the constant will not be necessary.

- Applying the generalized (weighted) least squares procedure to our household expenditure data yields the following estimates:

$$\hat{y}_t = 31.924 + 0.1410\ x_t \qquad (11.3.7)$$

$$(17.986)(0.0270)$$

- It is important to recognize that the interpretations for β_1 and β_2 are the same in the transformed model in (11.3.5) as they are in the untransformed model in (11.3.1).

The standard errors in (11.3.8), namely se($\hat{\beta}_1$) = 17.986 and se($\hat{\beta}_2$) = 0.0270 are both lower than their least squares counterparts that were calculated from White's estimator, namely se(b_1) = 23.704 and se(b_2) = 0.0382. Since generalized least squares is a better estimation procedure than least squares, we do expect the generalized least squares standard errors to be lower.

Remark: Remember that standard errors are square roots of *estimated* variances; in a single sample the relative magnitudes of variances may not always be reflected by their corresponding variance estimates. Thus, lower standard errors do not *always* mean better estimation.

- The smaller standard errors have the advantage of producing narrower more informative confidence intervals. For example, using the generalized least squares results, a 95% confidence interval for β_2 is given by

$$\hat{\hat{\beta}}_2 \pm t_c \mathrm{se}(\hat{\hat{\beta}}_2) = 0.1410 \pm 2.024(0.0270) = [0.086, 0.196]$$

The least squares confidence interval computed using White's standard errors was [0.051, 0.206].

11.4 Detecting Heteroskedasticity

11.4.1 Residual Plots

- One way of investigating the existence of heteroskedasticity is to estimate your model using least squares and to plot the least squares residuals.
- If the errors are homoskedastic, there should be no patterns of any sort in the residuals.

If the errors are heteroskedastic, they may tend to exhibit greater variation in some systematic way.

11.4.2 The Goldfeld-Quandt Test

- A formal test for heteroskedasticity is the Goldfeld-Quandt test. It involves the following steps:

1. Split the sample into two approximately equal subsamples. If heteroskedasticty exists, some observations will have large variances and others will have small variances.

Divide the sample such that the observations with potentially high variances are in one subsample and those with potentially low variances are in the other subsample.

2.Compute estimated error variances $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ for each of the subsamples. Let $\hat{\sigma}_1^2$ be the estimate from the subsample with potentially large variances and let $\hat{\sigma}_2^2$ be the estimate from the subsample with potentially small variances. If a null hypothesis of equal variances is not true, we expect $\hat{\sigma}_1^2/\hat{\sigma}_2^2$ to be large.

3.Compute $GQ = \hat{\sigma}_1^2/\hat{\sigma}_2^2$ and reject the null hypothesis of equal variances if $GQ > F_c$ where F_c is a critical value form the *F*-distribution with $(T_1 - K)$ and $(T_2 - K)$ degrees of freedom. The values T_1 and T_2 are the numbers of observations in each of the subsamples; if the sample is split exactly in half, $T_1 = T_2 = T/2$.

- Applying this test procedure to the household food expenditure model, we set up the hypotheses

$$H_0 : \sigma_t^2 = \sigma^2 \qquad H_1 : \sigma_t^2 = \sigma^2 x_t \tag{11.4.1}$$

- After ordering the data according to decreasing values of x_t, and using a partition of 20 observations in each subset of data, we find $\hat{\sigma}_1^2 = 2285.9$ and $\hat{\sigma}_2^2 = 682.46$. Hence, the value of the Goldfeld-Quandt statistic is

$$GQ = \frac{2285.9}{682.46} = 3.35$$

- The 5 percent critical value for (18, 18) degrees of freedom is $F_c = 2.22$. Thus, because $GQ = 3.35 > F_c = 2.22$, we reject H_0 and conclude that heteroskedasticity does exist; the error variance does depend on the level of income.

REMARK: The above test is a one-sided test because the alternative hypothesis suggested which sample partition will have the larger variance. If we suspect that two sample partitions could have different variances, but we do not know which variance is potentially larger,

11.5 A Sample With a Heteroskedastic Partition

11.5.1 Economic Model

- Consider modeling the supply of wheat in a particular wheat growing area in Australia. In the supply function the quantity of wheat supplied will typically depend upon the production technology of the firm, on the price of wheat or expectations about the price of wheat, and on weather conditions.

- We can depict this supply function as

$$\text{Quantity} = f(\text{Price, Technology, Weather}) \qquad (11.5.1)$$

- The data we have available from the Australian wheat growing district consist of 26 years of aggregate time-series data on quantity supplied and price.
- Because there is no obvious index of production technology, some kind of proxy needs to be used for this variable. We use a simple linear time-trend, a variable that takes the value 1 in year 1, 2 in year 2, and so on, up to 26 in year 26.
- An obvious weather variable is also unavailable; thus, in our statistical model, weather effects will form part of the random error term. Using these considerations, we specify the linear supply function

$$q_t = \beta_1 + \beta_2 p_t + \beta_3 t + e_t \qquad t = 1, 2, ..., 26 \qquad (11.5.2)$$

q_t is the quantity of wheat produced in year *t*,

p_t is the price of wheat guaranteed for year *t*,

$t = 1, 2, ..., 26$ is a trend variable introduced to capture changes in production technology, and

e_t is a random error term that includes, among other things, the influence of weather.

- To complete the econometric model in (11.5.2) some statistical assumptions for the random error term e_t are needed.
- In this case, however, we have additional information that makes an alternative assumption more realistic. After the 13th year, new wheat varieties whose yields are less susceptible to variations in weather conditions were introduced. These new varieties do not have an average yield that is higher than that of the old varieties, but the variance of their yields is lower because yield is less dependent on weather conditions.

- Since the weather effect is a major component of the random error term e_t, we can model the reduced weather effect of the last 13 years by assuming the error variance in those years is different from the error variance in the first 13 years. Thus, we assume that

$$
\begin{aligned}
&E(e_t) = 0 \\
&\operatorname{var}(e_t) = \sigma_1^2 \qquad t = 1, \ldots, 13 \qquad (11.5.3) \\
&\operatorname{var}(e_t) = \sigma_2^2 \qquad t = 14, \ldots, 26
\end{aligned}
$$

- From the above argument, we expect that $\sigma_2^2 < \sigma_1^2$.

11.5.2 Generalized Least Squares Through Model Transformation

- Write the model corresponding to the two subsets of observations as

$$
\begin{aligned}
q_t &= \beta_1 + \beta_2 p_t + \beta_3 t + e_t \qquad \operatorname{var}(e_t) = \sigma_1^2 \qquad t = 1, \ldots, 13 \\
q_t &= \beta_1 + \beta_2 p_t + \beta_3 t + e_t \qquad \operatorname{var}(e_t) = \sigma_2^2 \qquad t = 14, \ldots, 26
\end{aligned}
\qquad (11.5.4)
$$

- Dividing each variable by σ_1 for the first 13 observations and by σ_2 for the last 13 observations yields

$$\begin{aligned}
\frac{q_t}{\sigma_1} &= \beta_1 \frac{1}{\sigma_1} + \beta_2 \frac{p_t}{\sigma_1} + \beta_3 \frac{t}{\sigma_1} + \frac{e_t}{\sigma_1} \qquad t = 1, \ldots, 13 \\
\frac{q_t}{\sigma_2} &= \beta_1 \frac{1}{\sigma_2} + \beta_2 \frac{p_t}{\sigma_2} + \beta_3 \frac{t}{\sigma_2} + \frac{e_t}{\sigma_2} \qquad t = 14, \ldots, 26
\end{aligned} \tag{11.5.5}$$

- This transformation yields transformed error terms that have the same variance for all observations. Specifically, the transformed error variances are all equal to *one* because

$$\begin{aligned}
\operatorname{var}\left(\frac{e_t}{\sigma_1}\right) &= \frac{1}{\sigma_1^2} \operatorname{var}(e_t) = \frac{\sigma_1^2}{\sigma_1^2} = 1 \qquad t = 1, \ldots, 13 \\
\operatorname{var}\left(\frac{e_t}{\sigma_2}\right) &= \frac{1}{\sigma_2^2} \operatorname{var}(e_t) = \frac{\sigma_2^2}{\sigma_2^2} = 1 \qquad t = 14, \ldots, 26
\end{aligned}$$

- Providing σ_1 and σ_2 are known, the transformed model in (11.5.5) provides a set of new transformed variables to which we can apply the least squares principle to obtain the best linear unbiased estimator for (β_1, β_2, β_3).
- The transformed variables are

$$\frac{q_t}{\sigma_i} \quad \frac{1}{\sigma_i} \quad \frac{p_t}{\sigma_i} \quad \frac{t}{\sigma_i} \tag{11.5.6}$$

where σ_i is either σ_1 or σ_2, depending on which half of the observations are being considered.

- Like before, the complete process of transforming variables, then applying least squares to the transformed variables, is called generalized least squares.

11.5.3 Implementing Generalized Least Squares

- The transformed variables in (11.5.6) depend on the unknown variance parameters σ_1^2 and σ_2^2. Thus, as they stand, the transformed variables cannot be calculated.
- To overcome this difficulty, we use estimates of σ_1^2 and σ_2^2 and transform the variables as if the estimates were the true variances.
- It makes sense to split the sample into two, applying least squares to the first half to estimate σ_1^2 and applying least squares to the second half to estimate σ_2^2. Substituting these estimates for the true values causes no difficulties in large samples.
- For the wheat supply example we obtain

$$\hat{\sigma}_1^2 = 641.64 \qquad \hat{\sigma}_2^2 = 57.76 \tag{R11.7}$$

- Using these estimates to calculate observations on the transformed variables in (11.5.6), and then applying least squares to the complete sample defined in (11.5.5) yields the estimated equation:

$$\hat{q}_t = \underset{(12.7)}{138.1} + \underset{(8.81)}{21.72} p_t + \underset{(0.812)}{3.283} t \qquad \text{(R11.8)}$$

Remark: A word of warning about calculation of the standard errors is necessary. As demonstrated below (11.5.5), the transformed errors in (11.5.5) have a variance equal to *one*. However, when you transform your variables using $\hat{\sigma}_1$ and $\hat{\sigma}_2$, and apply least squares to the transformed variables for the complete sample, your computer program will automatically *estimate* a variance for the transformed errors. This estimate will not be *exactly* equal to *one*. The standard errors in (R11.8) were calculated by forcing the computer to use *one* as the variance of the transformed errors. Most software packages will have options that let you do this, but it is not crucial if your package does not; the variance estimate will usually be close to one anyway.

11.5.4 Testing the Variance Assumption

- To use a residual plot to check whether the wheat-supply error variance has decreased over time, it is sensible to plot the least-squares residuals against time. See Figure 11.3. The dramatic drop in the variation of the residuals after year 13 supports our belief that the variance has decreased.
- For the Goldfeld-Quandt test the sample is already split into two natural subsamples. Thus, we set up the hypotheses

$$H_0 : \sigma_1^2 = \sigma_2^2 \qquad H_1 : \sigma_2^2 < \sigma_1^2 \tag{11.5.9}$$

- The computed value of the Goldfeld-Quandt statistic is

$$GQ = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} = \frac{641.64}{57.76} = 11.11$$

- $T_1 = T_2 = 13$ and $K = 3$; thus, if H_0 is true, 11.11 is an observed value from an F-distribution with (10, 10) degrees of freedom. The corresponding 5 percent critical value is $F_c = 2.98$.
- Since $GQ = 11.11 > F_c = 2.98$, we reject H_0 and conclude that the observed difference between $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ could not reasonably be attributable to chance. There is evidence to suggest the new varieties have reduced the variance in the supply of wheat.