Chapter 11

Heteroskedasticity

11.1 The Nature of Heteroskedasticity

In Chapter 3 we introduced the linear model

\[ y = \beta_1 + \beta_2 x \]  \hspace{1cm} (11.1.1)

to explain household expenditure on food \((y)\) as a function of household income \((x)\). In this function \(\beta_1\) and \(\beta_2\) are unknown parameters that convey information about the expenditure function. The response parameter \(\beta_2\) describes how household food expenditure changes when household income increases by one unit. The intercept
parameter $\beta_1$ measures expenditure on food for a zero income level. Knowledge of these parameters aids planning by institutions such as government agencies or food retail chains.

- We begin this section by asking whether a function such as $y = \beta_1 + \beta_2x$ is better at explaining expenditure on food for low-income households than it is for high-income households.

- Low-income households do not have the option of extravagant food tastes; comparatively, they have few choices, and are almost forced to spend a particular portion of their income on food. High-income households, on the other hand, could have simple food tastes or extravagant food tastes. They might dine on caviar or spaghetti, while their low-income counterparts have to take the spaghetti.

- Thus, 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.
• To discover how, and what we mean by heteroskedasticity, let us return to the statistical model for the food expenditure-income relationship that we analysed in Chapters 3 through 6. Given $T = 40$ cross-sectional household observations on food expenditure and income, the statistical model specified in Chapter 3 was given by

$$y_t = \beta_1 + \beta_2 x_t + e_t$$  \hspace{1cm} (11.1.2)

where $y_t$ represents weekly food expenditure for the $t$-th household, $x_t$ represents weekly household income for the $t$-th household, and $\beta_1$ and $\beta_2$ are unknown parameters to estimate.

• Specifically, we assumed the $e_t$ were uncorrelated random error terms with mean zero and constant variance $\sigma^2$. That is,
\[ E(e_t) = 0 \quad \text{var}(e_t) = \sigma^2 \quad \text{cov}(e_i, e_j) = 0 \quad (11.1.3) \]

- Using the least squares procedure and the data in Table 3.1 we found estimates \( b_1 = 40.768 \) and \( b_2 = 0.1283 \) for the unknown parameters \( \beta_1 \) and \( \beta_2 \). Including the standard errors for \( b_1 \) and \( b_2 \), the estimated mean function was

\[
\hat{y}_t = 40.768 + 0.1283x_t
\]

(22.139) (0.0305)

- 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 points are scattered further away from the line as \( x_t \) gets larger.
• Another way to describe this feature is to say that the least squares residuals, defined by

\[ \hat{e}_t = y_t - b_1 - b_2 x_t \]  \hspace{1cm} (11.1.5)

increase in absolute value as income grows.

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

\[ e_t = y_t - \beta_1 - \beta_2 x_t \]  \hspace{1cm} (11.1.6)
Thus, the information in Figure 11.1 suggests that the unobservable errors also increase in absolute value as income ($x_t$) increases. That is, the variation of food expenditure $y_t$ around mean food expenditure $E(y_t)$ increases as income $x_t$ increases.

This observation is consistent with the hypothesis that we posed earlier, namely, that the mean food expenditure function is better at explaining food expenditure for low-income (spaghetti-eating) households than it is for high-income households who might be spaghetti eaters or caviar eaters.

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 $\sigma^2$. 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.

Instead, we should be looking for a way to model a variance $\sigma^2$ that increases as $x_t$ increases.
• Thus, we are questioning the constant variance assumption, which we have written as

\[
\text{var}(y_t) = \text{var}(e_t) = \sigma^2
\]  \hspace{1cm} (11.1.7)

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

\[
\text{var}(y_t) = \text{var}(e_t) = \sigma_t^2
\]  \hspace{1cm} (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 Equation (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. At $x_1$, the probability density function $f(y_1|x_1)$ is such that $y_1$ will be close to $E(y_1)$ with high probability. When we move to $x_2$, the probability density function $f(y_2|x_2)$ is more spread out; we are less certain about where $y_2$ might fall. When homoskedasticity exists, the probability density function for the errors does not change as $x$ changes, as we illustrated in Figure 3.3.

- The existence of different variances, or heteroskedasticity, is often encountered when using *cross-sectional data*. The term *cross-sectional data* refers to having data on a number of economic units such as firms or households, *at a given point in time*. The household data on income and food expenditure fall into this category.

- With *time-series data*, where we have data over time on one economic unit, such as a firm, a household, or even a whole economy, it is possible that the error variance will
change. This would be true if there was an external shock or change in circumstances that created more or less uncertainty about $y$.

- Given that we have a model that exhibits heteroskedasticity, we need to ask about the consequences on least squares estimation of the variation of one of our assumptions. Is there a better estimator that we can use? Also, how might we detect whether or not heteroskedasticity exists? It is to these questions that we now turn.
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 the best linear unbiased estimator (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.
- Now consider the following model

\[
y_t = \beta_1 + \beta_2 x_t + e_t
\]  

(11.2.1)

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

Note the heteroskedastic assumption \( \text{var}(e_t) = \sigma_t^2 \).

- In Chapter 4, Equation (4.2.1), we wrote the least squares estimator for \( \beta_2 \) as

\[ b_2 = \beta_2 + \sum w_t e_t \quad (11.2.2) \]

where

\[ w_t = \frac{x_t - \bar{x}}{\sum (x_t - \bar{x})^2} \]
This expression is a useful one for exploring the properties of least squares estimation under heteroskedasticity.

• The first property that we establish is that of unbiasedness. This property was derived under homoskedasticity in Equation (4.2.3) of Chapter 4. This proof still holds because the only error term assumption that it used, $E(e_t) = 0$, still holds. We reproduce it here for completeness.

\[
E(b_2) = E(\beta_2) + E(\Sigma w_i e_i) \\
= \beta_2 + \Sigma w_i E(e_i) = \beta_2
\]  

(11.2.4)

• The next result is that the least squares estimator is no longer best. That is, although it is still unbiased, it is no longer the best linear unbiased estimator. 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 $\text{var}(b_2)$ in Equation (4.2.11). From that equation, and using Equation (11.2.2), we have

$$\text{var}(b_2) = \text{var}(\beta_2) + \text{var}(\sum w_i e_i) = \text{var}(\sum w_i e_i)$$

$$= \sum w_i^2 \text{var}(e_i) + \sum_{i \neq j} \sum w_i w_j \text{cov}(e_i, e_j)$$

$$= \sum w_i^2 \sigma_i^2$$

$$= \frac{\sum \left[ (x_i - \bar{x})^2 \sigma_i^2 \right]}{\left[ \sum (x_i - \bar{x})^2 \right]^2}$$

(11.2.5)
In an earlier proof, where the variances were all the same \((\sigma_i^2 = \sigma^2)\), we were able to write the next-to-last line as \(\sigma_i^2 \sum w_i^2\). Now, the situation is more complex. Note from the last line in Equation (11.2.5) that

\[
\text{var}(b_2) \neq \frac{\sigma^2}{\sum (x_i - \bar{x})^2}
\]  

(11.2.6)

- Thus, if we use the least squares estimation procedure and ignore heteroskedasticity when it is present, we will be using an estimate of Equation (11.2.6) to obtain the standard error for \(b_2\), when in fact we should be using an estimate of Equation (11.2.5). Using incorrect standard errors means that interval estimates and hypothesis tests will no longer be valid. Note that standard computer software for least squares regression
will compute the estimated variance for $b_2$ based on Equation (11.2.6), unless told otherwise.

11.2.1 White’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 $\text{var}(b_2)$ is obtained by replacing $\sigma^2$ by the squares of the least squares residuals $\hat{e}_i^2$, in Equation (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
\[ \text{var}(b_1) = 561.89, \quad \text{var}(b_2) = 0.0014569 \]

Taking the square roots of these quantities yields the standard errors, so that we could write our estimated equation as

\[
\hat{y}_t = 40.768 + 0.1283x_t \\
(23.704) (0.0382) \quad \text{(White)} \\
(22.139) (0.0305) \quad \text{(incorrect)}
\]

- 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.
• Specifically, following Equation (5.1.12) of Chapter 5, we can construct two corresponding 95% confidence intervals for $\beta_2$.

  White: \[ b_2 \pm t_c \text{se}(b_2) = 0.1283 \pm 2.024(0.0382) = [0.051, 0.206] \]
  Incorrect: \[ b_2 \pm t_c \text{se}(b_2) = 0.1283 \pm 2.024(0.0305) = [0.067, 0.190] \]

If we ignore heteroskedasticity, we estimate that $\beta_2$ lie between 0.067 and 0.190. However, recognizing the existence of heteroskedasticity means recognizing that our information is less precise, and we estimate that $\beta_2$ lie between 0.051 and 0.206.

• White’s estimator for the standard errors helps overcome the problem of drawing incorrect inferences from least squares estimates in the presence of heteroskedasticity.

• However, if we can get a better estimator than least squares, then it makes more sense to use this better estimator and its corresponding standard errors. What is a “better
estimator” will depend on how we model the heteroskedasticity. That is, it will depend on what further assumptions we make about the $\sigma_i^2$. 
11.3 Proportional Heteroskedasticity

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

\[
y_t = \beta_1 + \beta_2 x_t + e_t
\]  

(11.3.1)

- Following the discussion in Section 11.1, we make the following assumptions:

\[
E(e_t) = 0, \quad \text{var}(e_t) = \sigma_i^2, \quad \text{cov}(e_i, e_j) = 0, \ (i \neq j)
\]
• By itself, the assumption \( \text{var}(e_t) = \sigma_t^2 \) is not adequate for developing a better procedure for estimating \( \beta_1 \) and \( \beta_2 \). We would need to estimate \( T \) different variances \( (\sigma_1^2, \sigma_2^2, \ldots, \sigma_T^2) \) plus \( \beta_1 \) and \( \beta_2 \), with only \( T \) sample observations; it is not possible to consistently estimate \( T \) or more parameters.

• We overcome this problem by making a further assumption about the \( \sigma_t^2 \). 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

\[
\text{var}(e_t) = \sigma_t^2 = \sigma^2 x_t
\]  

(11.3.2)

That is, we assume that the variance of the \( t \)-th error term \( \sigma_t^2 \) is given by a positive unknown constant parameter \( \sigma^2 \) multiplied by the positive income variable \( x_t \).
• As explained earlier, in economic terms this assumption implies that for low levels of income \( (x_t) \), food expenditure \( (y_t) \) will be clustered close to the mean function \( E(y_t) = \beta_1 + \beta_2 x_t \). Expenditure on food for low-income households will be largely explained by the level of income. At high levels of income, food expenditures can deviate more from the mean function. This means that there are likely to be many other factors, such as specified tastes and preferences, that reside in the error term, and that lead to a greater variation in food expenditure for high-income households.

• Thus, the assumption of heteroskedastic errors in Equation (11.3.2) is a reasonable one for the expenditure model.

• In any given practical setting it is important to think not only about whether the residuals from the data exhibit heteroskedasticity, but also about whether such heteroskedasticity is a likely phenomenon from an economic standpoint.

• 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.

- To demonstrate these facts, we begin by dividing both sides of the original equation in
  (11.3.1) by $\sqrt{x_t}$

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

(11.3.3)

Now, define the following transformed variables
The beauty of this transformed model is that the new transformed error term $e_t^*$ is homoskedastic. The proof of this result is:

$$\text{var}(e_t^*) = \text{var}\left(\frac{e_t}{\sqrt{x_t}}\right) = \frac{1}{x_t} \text{var}(e_t) = \frac{1}{x_t} \sigma^2 x_t = \sigma^2$$  \hspace{1cm} (11.3.6)
• The transformed error term will retain the properties \( E(e_i^*) = 0 \) and zero correlation between different observations, \( \text{cov}(e_i^*, e_j^*) = 0 \) 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 \( \beta_1 \) and \( \beta_2 \).

• Note that these transformed variables are all observable; it is a straightforward matter to compute “the observations” on these variables. Also, the transformed model is linear in the unknown parameters \( \beta_1 \) and \( \beta_2 \). These are the original parameters that we are interested in estimating. They have not been affected by the transformation.

• In short, the transformed model is a linear statistical model to which we can apply least squares estimation.

• 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.
To summarize, to obtain the best linear unbiased estimator for a model with heteroskedasticity of the type specified in Equation (11.3.2):

1. Calculate the transformed variables given in Equation (11.3.4).
2. Use least squares to estimate the transformed model given in Equation (11.3.5).

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 $\beta_1$ and $\beta_2$ that minimize the sum of squared errors. In this case, we are minimizing the sum of squared transformed errors that are given by

$$
\sum_{i=1}^{T} e_i^* = \sum_{i=1}^{T} \frac{e_i^2}{x_i}
$$
• 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:

\[ \hat{y}_t = 31.924 + 0.1410x_t \]  
\[(R11.4)\]

That is, we estimate the intercept term as \( \hat{\beta}_1 = 31.924 \) and the slope coefficient that shows that the response of food expenditure to a change in income as \( \hat{\beta}_2 = 0.1410 \). These estimates are somewhat different from the least squares estimate \( b_1 = 40.768 \) and \( b_2 = 0.1283 \) that did not allow for the existence of heteroskedasticity.

• It is important to recognize that the interpretations for \( \beta_1 \) and \( \beta_2 \) are the same in the transformed model in Equation (11.3.5) as they are in the untransformed model in Equation (11.3.1).
• Transformation of the variables should be regarded as a device for converting a heteroskedastic error model into a homoskedastic error model, not as something that changes the meaning of the coefficients.

• The standard errors in Equation (R11.4), 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 $\beta_2$ is given by

$$\hat{\beta}_2 \pm t_c \text{se}(\beta_2) = 0.1410 \pm 2.024(0.0270) = [0.086, 0.196]$$

The least squares confidence interval computed using White’s standard errors was $[0.051, 0.206]$. 
11.4 Detecting Heteroskedasticity

There is likely to be uncertainty about whether a heteroskedastic-error assumption is warranted. A common question is: How do we know if heteroskedasticity is likely to be a problem for our model and our set of data? Is there a way of detecting heteroskedasticity so that we know whether to use generalized least squares techniques? We will consider two ways of investigating these questions.

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.
- For example, for the household expenditure data, we suspected that the variance may increase as income increased. In Figure 11.1 we plotted the estimated least squares
function and the residuals and discovered that the absolute values of the residuals did indeed tend to increase as income increased. This method of investigating heteroskedasticity can be followed for any simple regression.

- When we have more than one explanatory variable, the estimated least squares function is not so easily depicted on a diagram. However, what we can do is plot the least squares residuals against each explanatory variable, against time, or against \( \hat{y}_t \), to see if those residuals vary in a systematic way relative to the specified variable.

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 heteroskedasticity 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. For example, in the food expenditure equation, where we believe the variances are related to $x_t$, the observations should be sorted according to the magnitude of $x_t$; the $T/2$ observations with the largest values of $x_t$ would form one subsample and the other $T/2$ observations, with the smallest values of $x_t$, would form the other.

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 from 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 as follows:

$$H_0 : \sigma_i^2 = \sigma^2 \quad H_1 : \sigma_i^2 = \sigma^2 x_i$$

(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, then a two-sided test with alternative hypothesis $H_1: \sigma_1^2 \neq \sigma_2^2$ is more appropriate. To perform a two-sided test at the 5 percent significance level we put the larger variance estimate in the numerator and use a critical value $F_c$ such that $P[F > F_c] = 0.025$. 
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

\[
\text{Quantity} = f(\text{Price}, \text{Technology}, \text{Weather})
\]  

(11.5.1)
• To estimate how the quantity supplied responds to price and other variables, we move from the economic model in Equation (11.5.1) to an econometric model that we can estimate.

• If we have a sample of time-series data, aggregated over all farms, there will be price variation from year to year, variation that can be used to estimate the response of quantity to price. Also, production technology will improve over time, meaning that a greater supply can become profitable at the same level of output price. Finally, a larger part of the year-to-year variation in supply could be attributable to weather conditions.

• 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

\[ q_t = \beta_1 + \beta_2 p_t + \beta_3 t + e_t \quad t = 1, 2, \ldots, 26 \]  \hspace{1cm} (11.5.2)

\( q_t \) is the quantity of wheat produced in year \( t \),
\( p_t \) is the price of wheat guaranteed for year \( t \),
\( t = 1, 2, \ldots, 26 \) 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.
As before, \( \beta_1, \beta_2, \) and \( \beta_3 \) are unknown parameters that we wish to estimate. The data
on \( q, p, \) and \( t \) are given in Table 11.1.

• To complete the econometric model in Equation (11.5.2) some statistical assumptions
for the random error term \( e_t \) are needed. One possibility is to assume the \( e_t \) are
independent identically distributed random variables with zero mean and constant variance. This assumption is in line with those made in earlier chapters.

- 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
\[ E(e_t) = 0 \quad t = 1, 2, \ldots, 26 \]
\[ \text{var}(e_t) = \sigma_1^2 \quad t = 1, 2, \ldots, 13 \]  \hspace{1cm} \tag{11.5.3}
\[ \text{var}(e_t) = \sigma_2^2 \quad t = 14, 15, \ldots, 26 \]
\[ \text{cov}(e_i, e_j) = 0 \quad i \neq j \]

From the above argument, we expect that \( \sigma_2^2 < \sigma_1^2 \).

- Since the error variance in Equation (11.5.3) is not constant for all observation, this model describes another form of heteroskedasticity. It is a form that partitions the sample into two subsets, one subset where the error variance is \( \sigma_1^2 \) and one where the error variance is \( \sigma_2^2 \).
11.5.2 Generalized Least Squares Through Model Transformation

Given the heteroskedastic error model with two variances, one for each subset of thirteen years, we consider transforming the model so that the variance of the transformed error term is constant over the whole sample. This approach made it possible to obtain a best linear unbiased estimator by applying least squares to the transformed model.

• Now we write the model corresponding to the two subsets of observations as

\[(11.5.4)\]
• Dividing each variable by $\sigma_1$ for the first 13 observations and by $\sigma_2$ for the last 13 observations yields

$$
\frac{q_t}{\sigma_1} = \beta_1 \frac{1}{\sigma_1} + \beta_2 \frac{p_t}{\sigma_1} + \beta_3 \frac{t}{\sigma_1} + \frac{e_t}{\sigma_1} \quad t = 1, \ldots, 13
$$

(11.5.5)

$$
\frac{q_t}{\sigma_2} = \beta_1 \frac{1}{\sigma_2} + \beta_2 \frac{p_t}{\sigma_2} + \beta_3 \frac{t}{\sigma_2} + \frac{e_t}{\sigma_2} \quad t = 14, \ldots, 26
$$

• 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
Providing $\sigma_1$ and $\sigma_2$ are known, the transformed model in Equation (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 $(\beta_1, \beta_2, \beta_3)$. The transformed variables are

\[
\begin{pmatrix}
\frac{q_t}{\sigma_i} \\
\frac{1}{\sigma_i} \\
\frac{p_t}{\sigma_i} \\
\frac{t}{\sigma_i}
\end{pmatrix}
\]  \hspace{1cm} (11.5.6)
where $\sigma_i$ is either $\sigma_1$ or $\sigma_2$, depending on which half of the observations are being considered.

- As 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 Equation (11.5.6) depend on the unknown variance parameters $\sigma_1^2$ and $\sigma_2^2$. Thus, as they stand, the transformed variables cannot be calculated. To overcome this difficulty, we use estimates of $\sigma_1^2$ and $\sigma_2^2$ and transform the variables as if the estimates were the true variances.

- Since $\sigma_1^2$ is the error variance from the first half of the sample and $\sigma_2^2$ is the error variance from the second half of the sample, it makes sense to split the sample into two, applying least squares to the first half to estimate $\sigma_1^2$ and applying least squares to
the second half to estimate $\sigma_2^2$. Substituting these estimates for the true values causes no difficulties in large samples.

- If we follow this strategy for the wheat supply example we obtain

\[ \sigma_1^2 = 641.64 \text{ and } \sigma_2^2 = 57.76 \]  \hspace{1cm} (R11.7)

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

\[ \hat{q}_t = 138.1 + 21.72 p_t + 3.283t \]  
\[ (12.7) \hspace{0.5cm} (8.81) \hspace{0.5cm} (0.812) \]  \hspace{1cm} (R11.8)
These estimates suggest that an increase in price of 1 unit will bring about an increase in supply of 21.72 units. The coefficient of the trend variable suggests that, each year, technological advances mean that an additional 3.283 units will be supplied, given constant prices.

The standard errors are sufficiently small to make the estimated coefficients significantly different from zero. However, the 95% confidence intervals for $\beta_2$ and $\beta_3$, derived using these standard errors, are relatively wide.

$$\beta_2 \pm t_c\text{se}(\beta_2) = 21.72 \pm 2.069(8.81) = [3.5, 39.9]$$

$$\beta_3 \pm t_c\text{se}(\beta_3) = 3.283 \pm 2.069(0.812) = [1.60, 4.96]$$
**Remark:** A word of warning about calculation of the standard errors is necessary. As demonstrated below Equation (11.5.5), the transformed errors in Equation (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 Equation (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

\[ H_0 : \sigma_1^2 = \sigma_2^2 \quad \text{and} \quad H_1 : \sigma_2^2 < \sigma_1^2 \]  

(11.5.7)

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.
<table>
<thead>
<tr>
<th>Exercise</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.1</td>
</tr>
<tr>
<td>11.9</td>
</tr>
</tbody>
</table>