We'll start with a quote from Studenmund:

The assumption of constant variances for different observations of the error term (homoskedasticity) is not always realistic. For example, in a model explaining heights, compare a one-inch error in measuring the height of a basketball player with a one-inch error in measuring the height of a mouse. It's likely that error term observations associated with the height of a basketball player would come from distributions with larger variances than those associated with the height of a mouse. As we'll see, the distinction between heteroskedasticity and homoskedasticity is important because OLS, when applied to heteroskedastic models, is no longer the minimum variance estimator (it is still unbiased, however).

So, the assumption of homoskedasticity says that the variance of the error term is exactly the same for each observation. To write this mathematically:

$$\text{VAR}(\varepsilon_i) = \sigma^2 \ (i=1,2,\ldots,n)$$

An assumption which may be more realistic than homoskedasticity is heteroskedasticity, that the variance of the error term may depend on the size of one of the explanatory variables. That is:

$$\text{VAR}(\varepsilon_i) = F(X_{1i}, X_{2i}, \ldots, X_{Ki}) \ (i=1,2,\ldots,n)$$

Examples

If you're explaining weight and using height as an explanatory variable, it seems like there would be a larger variance of the error term for taller people.

If you are doing a hedonic analysis of housing prices, the error term is likely to be larger for more expensive houses. While the value of the house is the dependent rather than an explanatory variable, it might be useful to model the error term as being related to, say, the square footage of the house.

If you're modeling personal income as a function of the level of education a person has received, you might think that the error term would be larger for those with more education. A person with a high school diploma may have fewer options to earn a really high income, while a person with an advanced degree may have more opportunities and, while not all of them will choose these options, the fact that some will should make for a greater variance in income among people with higher educational levels.
If you're looking at local pollution levels as a function of, say, neighborhood income levels, there might be a higher variance in pollution levels in low income neighborhoods than in high income neighborhoods.

**Consequences of Heteroskedasticity**
1. Heteroskedasticity does not cause bias in the coefficient estimates. As a result, the expected value of the estimates will be equal to the real value of the coefficients.
2. Heteroskedasticity increases the variances of the $\hat{\beta}$ distributions.
3. Heteroskedasticity causes OLS to underestimate the standard errors and overestimate the t-statistics of the estimated coefficients. As a result, estimated coefficients that are not significant may incorrectly appear to be significant.

**Testing for Heteroskedasticity**
There are a lot of ways to test for heteroskedasticity, but there's not really any way to prove that it exists. Basically, if you think it might be a problem, either from your own consideration of the problem or from previous research in the area, here are some ways to see if you should be addressing the problem.

1. Graphing residuals
Graph residuals against the explanatory variables and see if larger values of the explanatory variable are associated with larger residuals. Alternatively, you might calculate a correlation coefficient between each of the explanatory variables and the absolute values or squared values of the residuals. If there is a significant correlation, this could be a sign of heteroskedasticity.

2. Park test
Run your regression model and save the residuals. Then regress the residuals on the explanatory variable you think is related to the heteroskedasticity. Specifically, the Park test suggests the following functional form:

$$\ln(e_i^2) = \alpha_0 + \alpha_1\ln Z_i + \nu_i$$

where

- $e_i$ = the residual from the original regression for the $i$th observation
- $Z_i$ = the proportionality factor on which the heteroskedasticity is dependent. This may be one explanatory variable or a combination of explanatory variables.
- $\nu_i$ = a homoskedastic (with any luck) error term

If the estimated coefficient on $Z_i$ is significantly different from zero, this is evidence of heteroskedasticity.
3. Goldfeld-Quandt test
To use this test, you first need to sort your data by the explanatory variable or by the proportionality factor ($Z_i$) which you think is related to the heteroskedasticity. Then, omit the middle third of the observations (just temporarily) and do OLS regressions on the top third and on the bottom third of the observations. That is, estimate your model using only the observations with the largest values of $Z_i$ and then do a separate estimation of your model using only the observations with the smallest values of $Z_i$. Be sure that these two subsamples of the data have the same number of observations.

Calculate the value of the Goldfeld-Quandt test statistic:

$$GQ = \frac{RSS3}{RSS1}$$

where

- $RSS1$ is the residual sum of squares from the regression of your model on the first third of the data (with the smallest values of $Z_i$)
- $RSS3$ is the residual sum of squares from the regression of your model on the first third of the data (with the largest values of $Z_i$)

Then compare the value of $GQ$ with the critical value of the F-statistic to test the null hypothesis of homoskedasticity. The number of degrees of freedom for both the numerator and denominator is $n' - K - 1$, where $n'$ is equal to the number of observations in the regressions done using the top and bottom thirds of the data.

If $GQ$ is greater than the critical value then the residuals associated with the larger values of $Z_i$ are significantly greater than those associated with smaller values of $Z_i$ and the null hypothesis of homoskedasticity is rejected in favor of the alternative, heteroskedasticity.

4. White test
Basically, the White test involves regressing the squared residuals from the original regression on all the explanatory variables, and all the explanatory variables squared, and the products of all the pairs of the explanatory variables. The test statistic is $n*R^2$, which has a chi-square distribution with degrees of freedom equal to the number of slope coefficients in the second regression.

The original regression is run with the dependent and explanatory variables:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \epsilon_i$$

From this, the residuals are collected and the squared residuals are regressed on the explanatory variables, the explanatory variables squared and the products of all pairs of the explanatory variables:

$$(\epsilon_i)^2 = \alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_4 X_{1i}^2 + \alpha_5 X_{2i}^2 + \alpha_6 X_{3i}^2 + \alpha_7 X_{1i} X_{2i} + \alpha_8 X_{1i} X_{3i} + \alpha_9 X_{2i} X_{3i} + \nu_i$$
From this regression, multiply the number of observations (n) by the $R^2$ and compare it to the critical value from a Chi-square table where the number of degrees of freedom is equal to the number of slope coefficients in the second equation estimated (which would be 9 here).

**Remedies for Heteroskedasticity**

There are two options here. The first, weighted least squares, is to determine the factor on which the heteroskedasticity is based, the so-called proportionality factor (which Studenmund calls $Z$) and to divide each variable ($Y_i$ and $X_{1i}$ through $X_{Ki}$) of each observation by $Z_i$ and then do an OLS regression.

The second is to redefine the variables (perhaps to change the specification to some sort of semi-log model) to eliminate the heteroskedasticity.

There are no hard and fast rules for doing either of these. Basically, you need to figure out what the variance of the error terms or residuals depends on and make some sort of adjustment for it.

For example, if the standard error of the residuals seems to get larger as one of the explanatory variables increases in size you might use that variable as $Z$, the factor responsible for the heteroskedasticity.

1. **Weighted least squares, the brute force approach**

Weighted least squares is simply ordinary least squares where each observation is adjusted for the expected size of its error term. To do this, divide all the observations in each variable by whatever will make the error terms homoskedastic. For the sake of discussion, this is the proportionality factor, $Z_i$.

So, if the original equation to be estimated is

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$$

but, instead of

$$\text{VAR}(\varepsilon_i) = \text{VAR}(\varepsilon_i) = \sigma^2$$

we have pure heteroskedasticity, meaning that

$$\text{VAR}(\varepsilon_i) = s_i^2 = \sigma^2 Z_i^2$$

In this case, the original equation to be estimated can be rewritten as

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + Z_i \nu_i$$
where \( \nu_i \) is a proper homoskedastic error term. To isolate the homoskedastic error term, divide through by the proportionality factor, \( Z_i \):

\[
\frac{Y_i}{Z_i} = \frac{\beta_0}{Z_i} + \frac{\beta_1}{Z_i}X_{1i} + \frac{\beta_2}{Z_i}X_{2i} + \nu_i
\]

The error term, \( \nu_i \), is now homoskedastic.

The next step is to recalculate the values of the variables. Instead of \( X_{1i} \) and \( X_{2i} \), the new explanatory variables are \( \frac{\beta_0}{Z_i} \), \( \frac{X_{1i}}{Z_i} \) and \( \frac{X_{2i}}{Z_i} \). Instead of \( Y_i \), the dependent variable is \( \frac{Y_i}{Z_i} \).

How careful you need to be here depends on what the proportionality coefficient is.

Case 1: \( Z_i \) is not one of the original explanatory variables
In this case, there are now three explanatory variables in the equation and a constant needs to added:

\[
\frac{Y_i}{Z_i} = \alpha_0 + \frac{\beta_0}{Z_i} + \frac{\beta_1}{Z_i}X_{1i} + \frac{\beta_2}{Z_i}X_{2i} + \nu_i
\]

Case 2: \( Z_i \) is one of the original explanatory variables
If, for example, the proportionality coefficient, \( Z_i = X_{1i} \), then the second term on the right hand side of the rewritten equation can be expressed as:

\[
\frac{\beta_1}{Z_i}X_{1i} = \beta_1
\]

This term becomes the constant in the equation, so the new equation to be estimated is:

\[
\frac{Y_i}{Z_i} = \frac{\beta_0}{Z_i} + \beta_1 + \frac{\beta_2}{Z_i}X_{2i} + \nu_i
\]

As Studenmund says, you need to be very careful about interpreting these coefficients.

2. Redefining the variables, the Zen approach
This suggestion is difficult to discuss in general terms. Basically, if you have observations where one of the explanatory variables has a great range of sizes, this may lead to larger variances of the error term for larger values of the variable and smaller variances for smaller values of the variable.

Examples might include household income or state populations. Households with higher incomes might have larger variance in their level of, for example, consumption spending. Larger states may have higher variance of the number of people in prison.

A possible solution in each of these cases could be to divide the dependent variable by the explanatory variable whose wide range could affect the variance of the error terms.
Consumption spending might be divided by household income to become consumption as a proportion of income.

Prison population might be divided by total population to become prison population as a proportion of total population.