Since $\hat{y}$ is an estimate of the expected value of y, given the x_j, using (8.20) to test for heteroskedasticity is useful in cases where the variance is thought to change with the level of the expected value, $\mathrm{E}(y|\mathbf{x})$. The test from (8.20) can be viewed as a special case of the White test, since equation (8.20) can be shown to impose restrictions on the parameters in equation (8.19).

A Special Case of the White Test for Heteroskedasticity:

1. Estimate the model (8.10) by OLS, as usual. Obtain the OLS residuals $\hat{u}$ and the fitted values $\hat{y}$. Compute the squared OLS residuals $\hat{u}^2$ and the squared fitted values $\hat{y}^2$.
2. Run the regression in equation (8.20). Keep the R-squared from this regression, $R^2_{\hat{u}^2}$.
3. Form either the F or LM statistic and compute the p-value (using the $F_{2,n-3}$ distribution in the former case and the χ^2_2 distribution in the latter case).

EXAMPLE 8.5 SPECIAL FORM OF THE WHITE TEST IN THE LOG HOUSING PRICE EQUATION

We apply the special case of the White test to equation (8.18), where we use the LM form of the statistic. The important thing to remember is that the chi-square distribution always has two df. The regression of $\hat{u}^2$ on $\widehat{lprice}$, $(\widehat{lprice})^2$, where $\widehat{lprice}$ denotes the fitted values from (8.18), produces $R^2_{\hat{u}^2} = .0392$; thus, $LM = 88(.0392) \approx 3.45$, and the p-value $= .178$. This is stronger evidence of heteroskedasticity than is provided by the Breusch-Pagan test, but we still fail to reject homoskedasticity at even the 15% level.

Before leaving this section, we should discuss one important caveat. We have interpreted a rejection using one of the heteroskedasticity tests as evidence of heteroskedasticity. This is appropriate provided we maintain Assumptions MLR.1 through MLR.4. But, if MLR.4 is violated—in particular, if the functional form of $\mathrm{E}(y|\mathbf{x})$ is misspecified—then a test for heteroskedasticity can reject H_0, even if $\mathrm{Var}(y|\mathbf{x})$ is constant. For example, if we omit one or more quadratic terms in a regression model or use the level model when we should use the log, a test for heteroskedasticity can be significant. This has led some economists to view tests for heteroskedasticity as general misspecification tests. However, there are better, more direct tests for functional form misspecification, and we will cover some of them in Section 9.1. It is better to use explicit tests for functional form first, since functional form misspecification is more important than heteroskedasticity. Then, once we are satisfied with the functional form, we can test for heteroskedasticity.

8.4 Weighted Least Squares Estimation

If heteroskedasticity is detected using one of the tests in Section 8.3, we know from Section 8.2 that one possible response is to use heteroskedasticity-robust statistics after estimation by OLS. Before the development of heteroskedasticity-robust statistics, the response to a finding of heteroskedasticity was to specify its form and use a *weighted least squares* method, which we develop in this section. As we will argue, if we have correctly specified the form of the variance (as a function of explanatory variables), then weighted least squares (WLS) is more efficient than OLS, and WLS leads to new t and F statistics that have t and F distributions. We will also discuss the implications of using the wrong form of the variance in the WLS procedure.

The Heteroskedasticity Is Known up to a Multiplicative Constant

Let $\mathbf{x}$ denote all the explanatory variables in equation (8.10) and assume that

$$\text{Var}(u|\mathbf{x}) = \sigma^2 h(\mathbf{x}), \tag{8.21}$$

where $h(\mathbf{x})$ is some function of the explanatory variables that determines the heteroskedasticity. Since variances must be positive, $h(\mathbf{x}) > 0$ for all possible values of the independent variables. For now, we assume that the function $h(\mathbf{x})$ is known. The population parameter σ^2 is unknown, but we will be able to estimate it from a data sample.

For a random drawing from the population, we can write $\sigma_i^2 = \text{Var}(u_i|\mathbf{x}_i) = \sigma^2 h(\mathbf{x}_i) = \sigma^2 h_i$, where we again use the notation $\mathbf{x}_i$ to denote all independent variables for observation i, and h_i changes with each observation because the independent variables change across observations. For example, consider the simple savings function

$$sav_i = \beta_0 + \beta_1 inc_i + u_i \tag{8.22}$$

$$\text{Var}(u_i|inc_i) = \sigma^2 inc_i. \tag{8.23}$$

Here, $h(x) = h(inc) = inc$: the variance of the error is proportional to the level of income. This means that, as income increases, the variability in savings increases. (If $\beta_1 > 0$, the expected value of savings also increases with income.) Because inc is always positive, the variance in equation (8.23) is always guaranteed to be positive. The standard deviation of u_i, conditional on inc_i, is $\sigma\sqrt{inc_i}$.

How can we use the information in equation (8.21) to estimate the β_j? Essentially, we take the original equation,

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik} + u_i, \tag{8.24}$$

which contains heteroskedastic errors, and transform it into an equation that has homoskedastic errors (and satisfies the other Gauss-Markov assumptions). Since h_i is just a function of $\mathbf{x}_i$, $u_i/\sqrt{h_i}$ has a zero expected value conditional on $\mathbf{x}_i$. Further, since $\text{Var}(u_i|\mathbf{x}_i) = \text{E}(u_i^2|\mathbf{x}_i) = \sigma^2 h_i$, the variance of $u_i/\sqrt{h_i}$ (conditional on $\mathbf{x}_i$) is σ^2:

$$\text{E}\left((u_i/\sqrt{h_i})^2\right) = \text{E}(u_i^2)/h_i = (\sigma^2 h_i)/h_i = \sigma^2,$$

where we have suppressed the conditioning on $\mathbf{x}_i$ for simplicity. We can divide equation (8.24) by $\sqrt{h_i}$ to get

$$\begin{aligned} y_i/\sqrt{h_i} &= \beta_0/\sqrt{h_i} + \beta_1(x_{i1}/\sqrt{h_i}) + \beta_2(x_{i2}/\sqrt{h_i}) + \ldots \\ &\quad + \beta_k(x_{ik}/\sqrt{h_i}) + (u_i/\sqrt{h_i}) \end{aligned} \tag{8.25}$$

or

$$y_i^* = \beta_0 x_{i0}^* + \beta_1 x_{i1}^* + \ldots + \beta_k x_{ik}^* + u_i^*, \tag{8.26}$$

where $x_{i0}^* = 1/\sqrt{h_i}$ and the other starred variables denote the corresponding original variables divided by $\sqrt{h_i}$.

Equation (8.26) looks a little peculiar, but the important thing to remember is that we derived it so we could obtain estimators of the β_j that have better efficiency properties than OLS. The intercept β_0 in the original equation (8.24) is now multiplying the variable $x_{i0}^* = 1/\sqrt{h_i}$. Each slope parameter in β_j multiplies a new variable that rarely has a useful interpretation. This should not cause problems if we recall that, for interpreting the parameters and the model, we always want to return to the original equation (8.24).

In the preceding savings example, the transformed equation looks like

$$sav_i/\sqrt{inc_i} = \beta_0(1/\sqrt{inc_i}) + \beta_1\sqrt{inc_i} + u_i^*,$$

where we use the fact that $inc_i/\sqrt{inc_i} = \sqrt{inc_i}$. Nevertheless, β_1 is the marginal propensity to save out of income, an interpretation we obtain from equation (8.22).

Equation (8.26) is linear in its parameters (so it satisfies MLR.1), and the random sampling assumption has not changed. Further, u_i^* has a zero mean and a constant variance (σ^2), conditional on $\mathbf{x}_i^*$. This means that if the original equation satisfies the first four Gauss-Markov assumptions, then the transformed equation (8.26) satisfies all five Gauss-Markov assumptions. Also, if u_i has a normal distribution, then u_i^* has a normal distribution with variance σ^2. Therefore, the transformed equation satisfies the classical linear model assumptions (MLR.1 through MLR.6) if the original model does so except for the homoskedasticity assumption.

Since we know that OLS has appealing properties (is BLUE, for example) under the Gauss-Markov assumptions, the discussion in the previous paragraph suggests estimating the parameters in equation (8.26) by ordinary least squares. These estimators, β_0^*, β_1^*, …, β_k^*,will be different from the OLS estimators in the original equation. The β_j^* are examples of **generalized least squares (GLS) estimators**. In this case, the GLS estimators are used to account for heteroskedasticity in the errors. We will encounter other GLS estimators in Chapter 12.

Because equation (8.26) satisfies all of the ideal assumptions, standard errors, *t* statistics, and *F* statistics can all be obtained from regressions using the transformed variables. The sum of squared residuals from (8.26) divided by the degrees of freedom is an unbiased estimator of σ^2. Further, the GLS estimators, because they are the best linear unbiased estimators of the β_j, are necessarily more efficient than the OLS estimators $\hat{\beta}_j$ obtained from the untransformed equation. Essentially, after we have transformed the variables, we simply use standard OLS analysis. But we must remember to interpret the estimates in light of the original equation.

The *R*-squared that is obtained from estimating (8.26), while useful for computing *F* statistics, is not especially informative as a goodness-of-fit measure: it tells us how much variation in y^* is explained by the x_j^*, and this is seldom very meaningful.

The GLS estimators for correcting heteroskedasticity are called **weighted least squares (WLS) estimators**. This name comes from the fact that the β_j^* minimize the *weighted* sum of squared residuals, where each squared residual is weighted by $1/h_i$. The idea is that less weight is given to observations with a higher error variance; OLS gives each observation the same weight because it is best when the error variance is identical for all partitions of the population. Mathematically, the WLS estimators are the values of the b_j that make

$$\sum_{i=1}^{n} (y_i - b_0 - b_1 x_{i1} - b_2 x_{i2} - \dots - b_k x_{ik})^2 / h_i \quad \text{[8.27]}$$

as small as possible. Bringing the square root of $1/h_i$ inside the squared residual shows that the weighted sum of squared residuals is identical to the sum of squared residuals in the transformed variables:

$$\sum_{i=1}^{n} (y_i^* - b_0 x_{i0}^* - b_1 x_{i1}^* - b_2 x_{i2}^* - \dots - b_k x_{ik}^*)^2.$$

Since OLS minimizes the sum of squared residuals (regardless of the definitions of the dependent variable and independent variable), it follows that the WLS estimators that minimize (8.27) are simply the OLS estimators from (8.26). Note carefully that the squared residuals in (8.27) are weighted by $1/h_i$, whereas the transformed variables in (8.26) are weighted by $1/\sqrt{h_i}$.

A weighted least squares estimator can be defined for any set of positive weights. OLS is the special case that gives equal weight to all observations. The efficient procedure, GLS, weights each squared residual by the *inverse* of the conditional variance of u_i given $\mathbf{x}_i$.

Obtaining the transformed variables in equation (8.25) in order to manually perform weighted least squares can be tedious, and the chance of making mistakes is nontrivial. Fortunately, most modern regression packages have a feature for computing weighted least squares. Typically, along with the dependent and independent variables in the original model, we just specify the weighting function, $1/h_i$, appearing in (8.27). That is, we specify weights proportional to the inverse of the variance. In addition to making mistakes less likely, this forces us to interpret weighted least squares estimates in the original model. In fact, we can write out the estimated equation in the usual way. The estimates and standard errors will be different from OLS, but the way we *interpret* those estimates, standard errors, and test statistics is the same.

EXAMPLE 8.6 FINANCIAL WEALTH EQUATION

We now estimate equations that explain net total financial wealth (*nettfa*, measured in \$1,000s) in terms of income (*inc*, also measured in \$1,000s) and some other variables, including age, gender, and an indicator for whether the person is eligible for a 401(k) pension plan. We use the data on single people ($fsize = 1$) in 401KSUBS.RAW. In Computer Exercise C12 in Chapter 6, it was found that a specific quadratic function in *age*, namely $(age - 25)^2$, fit the data just as well as an unrestricted quadratic. Plus, the restricted form gives a simplified intepretation because the minimum age in the sample is 25: *nettfa* is an increasing function of *age* after $age = 25$.

The results are reported in Table 8.1. Because we suspect heteroskedasticity, we report the heteroskedasticity-robust standard errors for OLS. The weighted least squares estimates, and their standard errors, are obtained under the assumption $\text{Var}(u|inc) = \sigma^2 inc$.