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.

8-4a The Heteroskedasticity Is Known up to a Multiplicative Constant

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

$$\mathrm{Var}(u|\mathbf{x}) = \sigma^2 h(\mathbf{x}), \quad [8.21]$$

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 $\sigma_i^2 = \mathrm{Var}(u_i|\mathbf{x}_i) = \sigma^2 h(\mathbf{x}_i) = \sigma^2 h_i$, 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

$$sav_i = \beta_0 + \beta_1 inc_i + u_i \quad [8.22]$$

$$\mathrm{Var}(u_i|inc_i) = \sigma^2 inc_i. \quad [8.23]$$

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,

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

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:

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

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

$$y_i/\sqrt{h_i} = \beta_0/\sqrt{h_i} + \beta_1(x_{i1}/\sqrt{h_i}) + \beta_2(x_{i2}/\sqrt{h_i}) + \ldots + \beta_k(x_{ik}/\sqrt{h_i}) + (u_i/\sqrt{h_i}) \quad [8.25]$$

or

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

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, $\beta_0^*, \beta_1^*, \ldots, \beta_k^*$, 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 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

$$\sum_{i=1}^{n}(y_i - b_0 - b_1x_{i1} - b_2x_{i2} - \ldots - b_kx_{ik})^2/h_i \quad [8.27]$$

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:

$$\sum_{i=1}^{n}(y_i^* - b_0x_{i0}^* - b_1x_{i1}^* - b_2x_{i2}^* - \ldots - b_kx_{ik}^*)^2.$$

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.

Econometrics packages that have a built-in WLS option will report an R-squared (and adjusted R-squared) along with WLS estimates and standard errors. Typically, the WLS R-squared is obtained from the weighted SSR, obtained from minimizing equation (8.27), and a weighted total sum of squares (SST), obtained by using the same weights but setting all of the slope coefficients in equation (8.27), $b_1, b_2, \ldots, b_k$, to zero. As a goodness-of-fit measure, this R-squared is not especially useful, as it effectively measures explained variation in y_i^* rather than y_i. Nevertheless, the WLS R-squareds computed as just described are appropriate for computing F statistics for exclusion restrictions (provided we have properly specified the variance function). As in the case of OLS, the SST terms cancel, and so we obtain the F statistic based on the weighted SSR.

The R-squared from running the OLS regression in equation (8.26) is even less useful as a goodness-of-fit measure, as the computation of SST would make little sense: one would necessarily exclude an intercept from the regression, in which case regression packages typically compute the SST without properly centering the y_i^*. This is another reason for using a WLS option that is pre-programmed in a regression package because at least the reported R-squared properly compares the model with all of the independent variables to a model with only an intercept. Because the SST cancels out when testing exclusion restrictions, improperly computing SST does not affect the R-squared form of the F statistic. Nevertheless, computing such an R-squared tempts one to think the equation fits better than it does.

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. 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 interpretation 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 $\text{Var}(u|inc) = \sigma^2 inc$.

Without controlling for other factors, another dollar of income is estimated to increase *nettfa* by about 82¢ when OLS is used; the WLS estimate is smaller, about 79¢. The difference is not large; we certainly do not expect them to be identical. The WLS coefficient does have a smaller standard error than OLS, almost 40% smaller, provided we assume the model $\text{Var}(nettfa|inc) = \sigma^2 inc$ is correct.

Adding the other controls reduced the *inc* coefficient somewhat, with the OLS estimate still larger than the WLS estimate. Again, the WLS estimate of β_{inc} is more precise. Age has an increasing effect starting at $age = 25$, with the OLS estimate showing a larger effect. The WLS estimate of β_{age} is more precise in this case. Gender does not have a statistically significant effect on *nettfa*, but being eligible for a 401(k) plan does: the OLS estimate is that those eligible, holding fixed income, age, and gender, have net total financial assets about \$6,890 higher. The WLS estimate is substantially below the OLS estimate and suggests a misspecification of the functional form in the mean equation. (One possibility is to interact *e401k* and *inc*; see Computer Exercise C11.)

Using WLS, the *F* statistic for joint significance of $(age - 25)^2$, *male*, and *e401k* is about 30.8 if we use the *R*-squareds reported in Table 8.1. With 3 and 2,012 degrees of freedom, the *p*-value is zero to more than 15 decimal places; of course, this is not surprising given the very large *t* statistics for the age and 401(k) variables.

EXPLORING FURTHER 8.3

Using the OLS residuals obtained from the OLS regression reported in column (1) of Table 8.1, the regression of $\hat{u}^2$ on *inc* yields a *t* statistic of 2.96. Does it appear we should worry about heteroskedasticity in the financial wealth equation?

TABLE 8.1 Dependent Variable: *nettfa*

Independent Variables	(1) OLS	(2) WLS	(3) OLS	(4) WLS
inc	.821 (.104)	.787 (.063)	.771 (.100)	.740 (.064)
$(age - 25)^2$	—	—	.0251 (.0043)	.0175 (.0019)
male	—	—	2.48 (2.06)	1.84 (1.56)
e401k	—	—	6.89 (2.29)	5.19 (1.70)
intercept	−10.57 (2.53)	−9.58 (1.65)	−20.98 (3.50)	−16.70 (1.96)
Observations	2,017	2,017	2,017	2,017
R-squared	.0827	.0709	.1279	.1115

Assuming that the error variance in the financial wealth equation has a variance proportional to income is essentially arbitrary. In fact, in most cases, our choice of weights in WLS has a degree of arbitrariness. However, there is one case where the weights needed for WLS arise naturally from an underlying econometric model. This happens when, instead of using individual-level data, we only have averages of data across some group or geographic region. For example, suppose we are interested in determining the relationship between the amount a worker contributes to his or her 401(k) pension plan as a function of the plan generosity. Let i denote a particular firm and let e denote an employee within the firm. A simple model is

$$contrib_{i,e} = \beta_0 + \beta_1 earns_{i,e} + \beta_2 age_{i,e} + \beta_3 mrate_i + u_{i,e}, \tag{8.28}$$

where $contrib_{i,e}$ is the annual contribution by employee e who works for firm i, $earns_{i,e}$ is annual earnings for this person, and $age_{i,e}$ is the person's age. The variable $mrate_i$ is the amount the firm puts into an employee's account for every dollar the employee contributes.

If (8.28) satisfies the Gauss-Markov assumptions, then we could estimate it, given a sample on individuals across various employers. Suppose, however, that we only have *average* values of contributions, earnings, and age by employer. In other words, individual-level data are not available. Thus, let $\overline{contrib}_i$ denote average contribution for people at firm i, and similarly for $\overline{earns}_i$ and $\overline{age}_i$. Let m_i denote the number of employees at firm i; we assume that this is a known quantity. Then, if we average equation (8.28) across all employees at firm i, we obtain the firm-level equation

$$\overline{contrib}_i = \beta_0 + \beta_1 \overline{earns}_i + \beta_2 \overline{age}_i + \beta_3 mrate_i + \bar{u}_i, \tag{8.29}$$

where $\bar{u}_i = m_i^{-1}\sum_{e=1}^{mi} u_{i,e}$ is the average error across all employees in firm i. If we have n firms in our sample, then (8.29) is just a standard multiple linear regression model that can be estimated by OLS. The estimators are unbiased if the original model (8.28) satisfies the Gauss-Markov assumptions and the individual errors $u_{i,e}$ are independent of the firm's size, m_i [because then the expected value of $\bar{u}_i$, given the explanatory variables in (8.29), is zero].

If the individual-level equation (8.28) satisfies the homoskedasticity assumption, and the errors within firm i are uncorrelated across employees, then we can show that the firm-level equation (8.29) has a particular kind of heteroskedasticity. Specifically, if $\text{Var}(u_{i,e}) = \sigma^2$ for all i and e, and $\text{Cov}(u_{i,e}, u_{i,g}) = 0$ for every pair of employees $e \neq g$ within firm i, then $\text{Var}(\bar{u}_i) = \sigma^2/m_i$; this is just the usual formula for the variance of an average of uncorrelated random variables with common variance. In other words, the variance of the error term $\bar{u}_i$ decreases with firm size. In this case, $h_i = 1/m_i$, and so the most efficient procedure is weighted least squares, with weights equal to the number of employees at the firm $(1/h_i = m_i)$. This ensures that larger firms receive more weight. This gives us an efficient way of estimating the parameters in the individual-level model when we only have averages at the firm level.

A similar weighting arises when we are using per capita data at the city, county, state, or country level. If the individual-level equation satisfies the Gauss-Markov assumptions, then the error in the per capita equation has a variance proportional to one over the size of the population. Therefore, weighted least squares with weights equal to the population is appropriate. For example, suppose we have city-level data on per capita beer consumption (in ounces), the percentage of people in the population over 21 years old, average adult education levels, average income levels, and the city price of beer. Then, the city-level model

$$beerpc = \beta_0 + \beta_1 perc21 + \beta_2 avgeduc + \beta_3 incpc + \beta_4 price + u$$

can be estimated by weighted least squares, with the weights being the city population.

The advantage of weighting by firm size, city population, and so on relies on the underlying individual equation being homoskedastic. If heteroskedasticity exists at the individual level, then the proper weighting depends on the form of heteroskedasticity. Further, if there is correlation across errors within a group (say, firm), then $\text{Var}(\bar{u}_i) \neq \sigma^2/m_i$; see Problem 7. Uncertainty about the form of $\text{Var}(\bar{u}_i)$ in equations such as (8.29) is why more and more researchers simply use OLS and compute robust standard errors and test statistics when estimating models using per capita data. An alternative is to weight by group size but to report the heteroskedasticity-robust statistics in the WLS estimation. This ensures that, while the estimation is efficient if the individual-level model satisfies the Gauss-Markov assumptions, heteroskedasticity at the individual level or within-group correlation are accounted for through robust inference.

8-4b The Heteroskedasticity Function Must Be Estimated: Feasible GLS

In the previous subsection, we saw some examples of where the heteroskedasticity is known up to a multiplicative form. In most cases, the exact form of heteroskedasticity is not obvious. In other words, it is difficult to find the function $h(\mathbf{x}_i)$ of the previous section. Nevertheless, in many cases we can model the function h and use the data to estimate the unknown parameters in this model. This results in an estimate of each h_i, denoted as $\hat{h}_i$. Using $\hat{h}_i$ instead of h_i in the GLS transformation yields an estimator called the **feasible GLS (FGLS) estimator**. Feasible GLS is sometimes called *estimated GLS*, or EGLS.

There are many ways to model heteroskedasticity, but we will study one particular, fairly flexible approach. Assume that

$$\mathrm{Var}(u|\mathbf{x}) = \sigma^2\exp(\delta_0 + \delta_1 x_1 + \delta_2 x_2 + \ldots + \delta_k x_k), \qquad [8.30]$$

where $x_1, x_2, \ldots, x_k$ are the independent variables appearing in the regression model [see equation (8.1)], and the δ_j are unknown parameters. Other functions of the x_j can appear, but we will focus primarily on (8.30). In the notation of the previous subsection, $h(\mathbf{x}) = \exp(\delta_0 + \delta_1 x_1 + \delta_2 x_2 + \ldots + \delta_k x_k)$.

You may wonder why we have used the exponential function in (8.30). After all, when *testing* for heteroskedasticity using the Breusch-Pagan test, we assumed that heteroskedasticity was a linear function of the x_j. Linear alternatives such as (8.12) are fine when testing for heteroskedasticity, but they can be problematic when correcting for heteroskedasticity using weighted least squares. We have encountered the reason for this problem before: linear models do not ensure that predicted values are positive, and our estimated variances must be positive in order to perform WLS.

If the parameters δ_j were known, then we would just apply WLS, as in the previous subsection. This is not very realistic. It is better to use the data to estimate these parameters, and then to use these estimates to construct weights. How can we estimate the δ_j? Essentially, we will transform this equation into a linear form that, with slight modification, can be estimated by OLS.

Under assumption (8.30), we can write

$$u^2 = \sigma^2\exp(\delta_0 + \delta_1 x_1 + \delta_2 x_2 + \ldots + \delta_k x_k)v,$$

where v has a mean equal to unity, conditional on $\mathbf{x} = (x_1, x_2, \ldots, x_k)$. If we assume that v is actually independent of $\mathbf{x}$, we can write

$$\log(u^2) = \alpha_0 + \delta_1 x_1 + \delta_2 x_2 + \ldots + \delta_k x_k + e, \qquad [8.31]$$

where e has a zero mean and is independent of $\mathbf{x}$; the intercept in this equation is different from δ_0, but this is not important in implementing WLS. The dependent variable is the log of the squared error. Since (8.31) satisfies the Gauss-Markov assumptions, we can get unbiased estimators of the δ_j by using OLS.

As usual, we must replace the unobserved u with the OLS residuals. Therefore, we run the regression of

$$\log(\hat{u}^2) \text{ on } x_1, x_2, \ldots, x_k. \qquad [8.32]$$

Actually, what we need from this regression are the fitted values; call these $\hat{g}_i$. Then, the estimates of h_i are simply

$$\hat{h}_i = \exp(\hat{g}_i). \qquad [8.33]$$

We now use WLS with weights $1/\hat{h}_i$ in place of $1/h_i$ in equation (8.27). We summarize the steps.

A Feasible GLS Procedure to Correct for Heteroskedasticity:

1. Run the regression of y on $x_1, x_2, \ldots, x_k$ and obtain the residuals, $\hat{u}$.
2. Create $\log(\hat{u}^2)$ by first squaring the OLS residuals and then taking the natural log.
3. Run the regression in equation (8.32) and obtain the fitted values, $\hat{g}$.
4. Exponentiate the fitted values from (8.32): $\hat{h} = \exp(\hat{g})$.
5. Estimate the equation

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + u$$

by WLS, using weights $1/\hat{h}$. In other words, we replace h_i with $\hat{h}_i$ in equation (8.27). Remember, the *squared* residual for observation i gets weighted by $1/\hat{h}_i$. If instead we first transform all variables and run OLS, each variable gets multiplied by $1/\sqrt{\hat{h}_i}$, including the intercept.

If we could use h_i rather than $\hat{h}_i$ in the WLS procedure, we know that our estimators would be unbiased; in fact, they would be the best linear unbiased estimators, assuming that we have properly modeled the heteroskedasticity. Having to estimate h_i using the same data means that the FGLS estimator is no longer unbiased (so it cannot be BLUE, either). Nevertheless, the FGLS estimator is consistent and *asymptotically* more efficient than OLS. This is difficult to show because of estimation of the variance parameters. But if we ignore this—as it turns out we may—the proof is similar to showing that OLS is efficient in the class of estimators in Theorem 5.3. At any rate, for large sample sizes, FGLS is an attractive alternative to OLS when there is evidence of heteroskedasticity that inflates the standard errors of the OLS estimates.

We must remember that the FGLS estimators are estimators of the parameters in the usual population model

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + u.$$

Just as the OLS estimates measure the marginal impact of each x_j on y, so do the FGLS estimates. We use the FGLS estimates in place of the OLS estimates because the FGLS estimators are more efficient and have associated test statistics with the usual t and F distributions, at least in large samples. If we have some doubt about the variance specified in equation (8.30), we can use heteroskedasticity-robust standard errors and test statistics in the transformed equation.

Another useful alternative for estimating h_i is to replace the independent variables in regression (8.32) with the OLS fitted values and their squares. In other words, obtain the $\hat{g}_i$ as the fitted values from the regression of

$$\log(\hat{u}^2) \text{ on } \hat{y}, \hat{y}^2 \qquad [8.34]$$

and then obtain the $\hat{h}_i$ exactly as in equation (8.33). This changes only step (3) in the previous procedure.

If we use regression (8.32) to estimate the variance function, you may be wondering if we can simply test for heteroskedasticity using this same regression (an F or LM test can be used). In fact, Park (1966) suggested this. Unfortunately, when compared with the tests discussed in Section 8-3, the Park test has some problems. First, the null hypothesis must be something stronger than homoskedasticity: effectively, u and $\mathbf{x}$ must be independent. This is not required in the Breusch-Pagan or White tests. Second, using the OLS residuals $\hat{u}$ in place of u in (8.32) can cause the F statistic to deviate from the F distribution, even in large sample sizes. This is not an issue in the other tests we have covered. For these reasons, the Park test is not recommended when testing for heteroskedasticity. Regression (8.32) works well for weighted least squares because we only need consistent estimators of the δ_j, and regression (8.32) certainly delivers those.

EXAMPLE 8.7 Demand for Cigarettes

We use the data in SMOKE to estimate a demand function for daily cigarette consumption. Since most people do not smoke, the dependent variable, *cigs*, is zero for most observations. A linear model is not ideal because it can result in negative predicted values. Nevertheless, we can still learn something about the determinants of cigarette smoking by using a linear model.

The equation estimated by ordinary least squares, with the usual OLS standard errors in parentheses, is

$$\begin{aligned}\widehat{cigs} = &-3.64 + .880 \log(income) - .751 \log(cigpric)\\ &(24.08) \quad (.728) \qquad\qquad (5.773)\\ &- .501\, educ + .771\, age - .0090\, age^2 - 2.83\, restaurn\\ &\quad (.167) \qquad (.160) \qquad (.0017) \qquad (1.11)\\ n = &\ 807, R^2 = .0526,\end{aligned} \tag{8.35}$$

where

cigs = number of cigarettes smoked per day.
income = annual income.
cigpric = the per-pack price of cigarettes (in cents).
educ = years of schooling.
age = age measured in years.
restaurn = a binary indicator equal to unity if the person resides in a state with restaurant smoking restrictions.

Since we are also going to do weighted least squares, we do not report the heteroskedasticity-robust standard errors for OLS. (Incidentally, 13 out of the 807 fitted values are less than zero; this is less than 2% of the sample and is not a major cause for concern.)

Neither income nor cigarette price is statistically significant in (8.35), and their effects are not practically large. For example, if income increases by 10%, *cigs* is predicted to increase by $(.880/100)(10) = .088$, or less than one-tenth of a cigarette per day. The magnitude of the price effect is similar.

Each year of education reduces the average cigarettes smoked per day by one-half of a cigarette, and the effect is statistically significant. Cigarette smoking is also related to age, in a quadratic fashion. Smoking increases with age up until $age = .771/[2(.009)] \approx 42.83$, and then smoking decreases with age. Both terms in the quadratic are statistically significant. The presence of a restriction on smoking in restaurants decreases cigarette smoking by almost three cigarettes per day, on average.

Do the errors underlying equation (8.35) contain heteroskedasticity? The Breusch-Pagan regression of the squared OLS residuals on the independent variables in (8.35) [see equation (8.14)] produces $R^2_{\hat{u}^2} = .040$. This small *R*-squared may seem to indicate no heteroskedasticity, but we must remember to compute either the *F* or *LM* statistic. If the sample size is large, a seemingly small $R^2_{\hat{u}^2}$ can result in a very strong rejection of homoskedasticity. The *LM* statistic is $LM = 807(.040) = 32.28$, and this is the outcome of a χ^2_6 random variable. The *p*-value is less than .000015, which is very strong evidence of heteroskedasticity.

Therefore, we estimate the equation using the feasible GLS procedure based on equation (8.32). The weighted least squares estimates are

$$\begin{aligned}\widehat{cigs} = &\ 5.64 + 1.30 \log(income) - 2.94 \log(cigpric)\\ &(17.80) \quad (.44) \qquad\qquad (4.46)\\ &- .463\, educ + .482\, age - .0056\, age^2 - 3.46\, restaurn\\ &\quad (.120) \qquad (.097) \qquad (.0009) \qquad (.80)\\ n = &\ 807, R^2 = .1134.\end{aligned} \tag{8.36}$$

The income effect is now statistically significant and larger in magnitude. The price effect is also notably bigger, but it is still statistically insignificant. [One reason for this is that *cigpric* varies only across states in the sample, and so there is much less variation in log(*cigpric*) than in log(*income*), *educ*, and *age*.]

The estimates on the other variables have, naturally, changed somewhat, but the basic story is still the same. Cigarette smoking is negatively related to schooling, has a quadratic relationship with *age*, and is negatively affected by restaurant smoking restrictions.

We must be a little careful in computing *F* statistics for testing multiple hypotheses after estimation by WLS. (This is true whether the sum of squared residuals or *R*-squared form of the *F* statistic is used.) It is important that the same weights be used to estimate the unrestricted and restricted models. We should first estimate the unrestricted model by OLS. Once we have obtained the weights, we can use them to estimate the restricted model as well. The *F* statistic can be computed as usual. Fortunately, many regression packages have a simple command for testing joint restrictions after WLS estimation, so we need not perform the restricted regression ourselves.

Example 8.7 hints at an issue that sometimes arises in applications of weighted least squares: the OLS and WLS estimates can be substantially different. This is not such a big problem in the demand for cigarettes equation because all the coefficients maintain the same signs, and the biggest changes are on variables that were statistically insignificant when the equation was estimated by OLS. The OLS and WLS estimates will always differ due to sampling error. The issue is whether their difference is enough to change important conclusions.

If OLS and WLS produce statistically significant estimates that differ in sign—for example, the OLS price elasticity is positive and significant, while the WLS price elasticity is negative and significant—or the difference in magnitudes of the estimates is practically large, we should be suspicious. Typically, this indicates that one of the *other* Gauss-Markov assumptions is false, particularly the zero conditional mean assumption on the error (MLR.4). If $\mathrm{E}(y|\mathbf{x}) \neq \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k$, then OLS and WLS have different expected values and probability limits. For WLS to be consistent for the β_j, it is not enough for u to be uncorrelated with each x_j; we need the stronger assumption MLR.4 in the linear model MLR.1. Therefore, a significant difference between OLS and WLS can indicate a functional form misspecification in $\mathrm{E}(y|\mathbf{x})$. The *Hausman test* [Hausman (1978)] can be used to formally compare the OLS and WLS estimates to see if they differ by more than sampling error suggests they should, but this test is beyond the scope of this text. In many cases, an informal "eyeballing" of the estimates is sufficient to detect a problem.

EXPLORING FURTHER 8.4

Let $\hat{u}_i$ be the WLS residuals from (8.36), which are not weighted, and let $\widehat{cigs}_i$ be the fitted values. (These are obtained using the same formulas as OLS; they differ because of different estimates of the β_j.) One way to determine whether heteroskedasticity has been eliminated is to use the $\hat{u}_i^2/\hat{h}_i = (\hat{u}_i/\sqrt{\hat{h}_i})^2$ in a test for heteroskedasticity. [If $h_i = \mathrm{Var}(u_i|\mathbf{x}_i)$, then the transformed residuals should have little evidence of heteroskedasticity.] There are many possibilities, but one—based on White's test in the transformed equation—is to regress $\hat{u}_i^2/\hat{h}_i$ on $\widehat{cigs}_i/\sqrt{\hat{h}_i}$ and $\widehat{cigs}_i^2/\hat{h}_i$ (including an intercept). The joint F statistic when we use SMOKE is 11.15. Does it appear that our correction for heteroskedasticity has actually eliminated the heteroskedasticity?

8-4c What If the Assumed Heteroskedasticity Function Is Wrong?

We just noted that if OLS and WLS produce very different estimates, it is likely that the conditional mean $\mathrm{E}(y|\mathbf{x})$ is misspecified. What are the properties of WLS if the variance function we use is misspecified in the sense that $\mathrm{Var}(y|\mathbf{x}) \neq \sigma^2 h(\mathbf{x})$ for our chosen function $h(\mathbf{x})$? The most important issue is whether misspecification of $h(\mathbf{x})$ causes bias or inconsistency in the WLS estimator. Fortunately, the answer is no, at least under MLR.4. Recall that, if $E(u|\mathbf{x}) = 0$, then any function of $\mathbf{x}$ is uncorrelated with u, and so the weighted error, $u/\sqrt{h(\mathbf{x})}$, is uncorrelated with the weighted regressors, $x_j/\sqrt{h(\mathbf{x})}$, for *any* function $h(\mathbf{x})$ that is always positive. This is why, as we just discussed, we can take large differences between the OLS and WLS estimators as indicative of functional form misspecification. If we estimate parameters in the function, say $h(\mathbf{x}, \hat{\delta})$, then we can no longer claim that WLS is unbiased, but it will generally be consistent (whether or not the variance function is correctly specified).

If WLS is at least consistent under MLR.1 to MLR.4, what are the consequences of using WLS with a misspecified variance function? There are two. The first, which is very important, is that the usual WLS standard errors and test statistics, computed under the assumption that $\text{Var}(y|\mathbf{x}) = \sigma^2 h(\mathbf{x})$, are no longer valid, even in large samples. For example, the WLS estimates and standard errors in column (4) of Table 8.1 assume that $\text{Var}(nettfa|inc, age, male, e401k) = \text{Var}(nettfa|inc) = \sigma^2 inc$; so we are assuming not only that the variance depends just on income, but also that it is a linear function of income. If this assumption is false, the standard errors (and any statistics we obtain using those standard errors) are not valid. Fortunately, there is an easy fix: just as we can obtain standard errors for the OLS estimates that are robust to arbitrary heteroskedasticity, we can obtain standard errors for WLS that allow the variance function to be arbitrarily misspecified. It is easy to see why this works. Write the transformed equation as

$$y_i/\sqrt{h_i} = \beta_0(1/\sqrt{h_i}) + \beta_1(x_{i1}/\sqrt{h_i}) + \ldots + \beta_k(x_{ik}/\sqrt{h_i}) + u_i/\sqrt{h_i}.$$

Now, if $\text{Var}(u_i|\mathbf{x}_i) \neq \sigma^2 h_i$, then the weighted error $u_i/\sqrt{h_i}$ is heteroskedastic. So we can just apply the usual heteroskedasticity-robust standard errors after estimating this equation by OLS—which, remember, is identical to WLS.

To see how robust inference with WLS works in practice, column (1) of Table 8.2 reproduces the last column of Table 8.1, and column (2) contains standard errors robust to $\text{Var}(u_i|\mathbf{x}_i) \neq \sigma^2 inc_i$.

The standard errors in column (2) allow the variance function to be misspecified. We see that, for the income and age variables, the robust standard errors are somewhat above the usual WLS standard errors—certainly by enough to stretch the confidence intervals. On the other hand, the robust standard errors for *male* and *e401k* are actually smaller than those that assume a correct variance function. We saw this could happen with the heteroskedasticity-robust standard errors for OLS, too.

Even if we use flexible forms of variance functions, such as that in (8.30), there is no guarantee that we have the correct model. While exponential heteroskedasticity is appealing and reasonably flexible, it is, after all, just a model. Therefore, it is always a good idea to compute fully robust standard errors and test statistics after WLS estimation.

TABLE 8.2 WLS Estimation of the *nettfa* Equation

Independent Variables	With Nonrobust Standard Errors	With Robust Standard Errors
inc	.740 (.064)	.740 (.075)
$(age - 25)^2$	.0175 (.0019)	.0175 (.0026)
male	1.84 (1.56)	1.84 (1.31)
e401k	5.19 (1.70)	5.19 (1.57)
intercept	−16.70 (1.96)	−16.70 (2.24)
Observations	2,017	2,017
R-squared	.1115	.1115

A modern criticism of WLS is that if the variance function is misspecified, it is not guaranteed to be more efficient than OLS. In fact, that is the case: if $\text{Var}(y|\mathbf{x})$ is neither constant nor equal to $\sigma^2 h(\mathbf{x})$, where $h(\mathbf{x})$ is the proposed model of heteroskedasticity, then we cannot rank OLS and WLS in terms of variances (or asymptotic variances when the variance parameters must be estimated). However, this theoretically correct criticism misses an important practical point. Namely, in cases of strong heteroskedasticity, it is often better to use a wrong form of heteroskedasticity and apply WLS than to ignore heteroskedasticity altogether in estimation and use OLS. Models such as (8.30) can well approximate a variety of heteroskedasticity functions and may produce estimators with smaller (asymptotic) variances. Even in Example 8.6, where the form of heteroskedasticity was assumed to have the simple form $\text{Var}(nettfa|\mathbf{x}) = \sigma^2 inc$, the fully robust standard errors for WLS are well below the fully robust standard errors for OLS. (Comparing robust standard errors for the two estimators puts them on equal footing: we assume neither homoskedasticity nor that the variance has the form $\sigma^2 inc$.) For example, the robust standard error for the WLS estimator of β_{inc} is about .075, which is 25% lower than the robust standard error for OLS (about .100). For the coefficient on $(age - 25)^2$, the robust standard error of WLS is about .0026, almost 40% below the robust standard error for OLS (about .0043).

8-4d Prediction and Prediction Intervals with Heteroskedasticity

If we start with the standard linear model under MLR.1 to MLR.4, but allow for heteroskedasticity of the form $\text{Var}(y|\mathbf{x}) = \sigma^2 h(\mathbf{x})$ [see equation (8.21)], the presence of heteroskedasticity affects the point prediction of y only insofar as it affects estimation of the β_j. Of course, it is natural to use WLS on a sample of size n to obtain the $\hat{\beta}_j$. Our prediction of an unobserved outcome, y^0, given known values of the explanatory variables $\mathbf{x}^0$, has the same form as in Section 6-4: $\hat{y}^0 = \hat{\beta}_0 + \mathbf{x}^0\hat{\boldsymbol{\beta}}$. This makes sense: once we know $\text{E}(y|\mathbf{x})$, we base our prediction on it; the structure of $\text{Var}(y|\mathbf{x})$ plays no direct role.

On the other hand, prediction *intervals* do depend directly on the nature of $\text{Var}(y|\mathbf{x})$. Recall in Section 6-4 that we constructed a prediction interval under the classical linear model assumptions. Suppose now that all the CLM assumptions hold except that (8.21) replaces the homoskedasticity assumption, MLR.5. We know that the WLS estimators are BLUE and, because of normality, have (conditional) normal distributions. We can obtain $\text{se}(\hat{y}^0)$ using the same method in Section 6-4, except that now we use WLS. [A simple approach is to write $y_i = \theta_0 + \beta_1(x_{i1} - x_1^0) + \ldots + \beta_k(x_{ik} - x_k^0) + u_i$, where the x_j^0 are the values of the explanatory variables for which we want a predicted value of y. We can estimate this equation by WLS and then obtain $\hat{y}^0 = \hat{\theta}_0$ and $\text{se}(\hat{y}^0) = \text{se}(\hat{\theta}_0)$.] We also need to estimate the standard deviation of u^0, the unobserved part of y^0. But $\text{Var}(u^0|\mathbf{x} = \mathbf{x}^0) = \sigma^2 h(\mathbf{x}^0)$, and so $\text{se}(u^0) = \hat{\sigma}\sqrt{h(\mathbf{x}^0)}$, where $\hat{\sigma}$ is the standard error of the regression from the WLS estimation. Therefore, a 95% prediction interval is

$$\hat{y}^0 \pm t_{.025} \cdot \text{se}(\hat{e}^0), \qquad [8.37]$$

where $\text{se}(\hat{e}^0) = \{[\text{se}(\hat{y}^0)]^2 + \hat{\sigma}^2 h(\mathbf{x}^0)\}^{1/2}$.

This interval is exact only if we do not have to estimate the variance function. If we estimate parameters, as in model (8.30), then we cannot obtain an exact interval. In fact, accounting for the estimation error in the $\hat{\beta}_j$ and the $\hat{\delta}_j$ (the variance parameters) becomes very difficult. We saw two examples in Section 6-4 where the estimation error in the parameters was swamped by the variation in the unobservables, u^0. Therefore, we might still use equation (8.37) with $h(\mathbf{x}^0)$ simply replaced by $\hat{h}(\mathbf{x}^0)$. In fact, if we are to ignore the parameter estimation error entirely, we can drop $\text{se}(\hat{y}^0)$ from $\text{se}(\hat{e}^0)$. [Remember, $\text{se}(\hat{y}^0)$ converges to zero at the rate $1/\sqrt{n}$, while $\text{se}(\hat{u}^0)$ is roughly constant.]

We can also obtain a prediction for y in the model

$$\log(y) = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + u, \qquad [8.38]$$

where u is heteroskedastic. We assume that u has a conditional normal distribution with a specific form of heteroskedasticity. We assume the exponential form in equation (8.30), but add the normality assumption:

$$u|x_1, x_2, \ldots, x_k \sim \text{Normal}[0, \exp(\delta_0 + \delta_1 x_1 + \ldots + \delta_k x_k)]. \quad [8.39]$$

As a notational shorthand, write the variance function as $\exp(\delta_0 + \mathbf{x}\boldsymbol{\delta})$. Then, because $\log(y)$ given $\mathbf{x}$ has a normal distribution with mean $\beta_0 + \mathbf{x}\boldsymbol{\beta}$ and variance $\exp(\delta_0 + \mathbf{x}\boldsymbol{\delta})$, it follows that

$$\text{E}(y|\mathbf{x}) = \exp(\beta_0 + \mathbf{x}\boldsymbol{\beta} + \exp(\delta_0 + \mathbf{x}\boldsymbol{\delta})/2). \quad [8.40]$$

Now, we estimate the β_j and δ_j using WLS estimation of (8.38). That is, after using OLS to obtain the residuals, run the regression in (8.32) to obtain fitted values,

$$\hat{g}_i = \hat{\alpha}_0 + \hat{\delta}_1 x_{i1} + \ldots + \hat{\delta}_k x_{ik}, \quad [8.41]$$

and then compute the $\hat{h}_i$ as in (8.33). Using these $\hat{h}_i$, obtain the WLS estimates, $\hat{\beta}_j$, and also compute $\hat{\sigma}^2$ from the weighted squared residuals. Now, compared with the original model for $\text{Var}(u|\mathbf{x})$, $\delta_0 = \alpha_0 + \log(\sigma^2)$, and so $\text{Var}(u|\mathbf{x}) = \sigma^2 \exp(\alpha_0 + \delta_1 x_1 + \ldots + \delta_k x_k)$. Therefore, the estimated variance is $\hat{\sigma}^2 \exp(\hat{g}_i) = \hat{\sigma}^2 \hat{h}_i$, and the fitted value for y_i is

$$\hat{y}_i = \exp(\widehat{logy}_i + \hat{\sigma}^2 \hat{h}_i/2). \quad [8.42]$$

We can use these fitted values to obtain an R-squared measure, as described in Section 6-4: use the squared correlation coefficient between y_i and $\hat{y}_i$.

For any values of the explanatory variables $\mathbf{x}^0$, we can estimate $\text{E}(y|\mathbf{x} = \mathbf{x}^0)$ as

$$\hat{\text{E}}(y|\mathbf{x} = \mathbf{x}^0) = \exp(\hat{\beta}_0 + \mathbf{x}^0\hat{\boldsymbol{\beta}} + \hat{\sigma}^2 \exp(\hat{\alpha}_0 + \mathbf{x}^0\hat{\boldsymbol{\delta}})/2), \quad [8.43]$$

where

$\hat{\beta}_j$ = the WLS estimates.
$\hat{\alpha}_0$ = the intercept in (8.41).
$\hat{\delta}_j$ = the slopes from the same regression.
$\hat{\sigma}^2$ is obtained from the WLS estimation.

Obtaining a proper standard error for the prediction in (8.42) is very complicated analytically, but, as in Section 6-4, it would be fairly easy to obtain a standard error using a resampling method such as the bootstrap described in Appendix 6A.

Obtaining a prediction interval is more of a challenge when we estimate a model for heteroskedasticity, and a full treatment is complicated. Nevertheless, we saw in Section 6-4 two examples where the error variance swamps the estimation error, and we would make only a small mistake by ignoring the estimation error in all parameters. Using arguments similar to those in Section 6-4, an approximate 95% prediction interval (for large sample sizes) is $\exp[-1.96 \cdot \hat{\sigma}\sqrt{\hat{h}(\mathbf{x}^0)}]\exp(\hat{\beta}_0 + \mathbf{x}^0\hat{\boldsymbol{\beta}})$ to $\exp[1.96 \cdot \hat{\sigma}\sqrt{\hat{h}(\mathbf{x}^0)}]\exp(\hat{\beta}_0 + \mathbf{x}^0\hat{\boldsymbol{\beta}})$, where $\hat{h}(\mathbf{x}^0)$ is the estimated variance function evaluated at $\mathbf{x}^0$, $\hat{h}(\mathbf{x}^0) = \exp(\hat{\alpha}_0 + \hat{\delta}_1 x_1^0 + \ldots + \hat{\delta}_k x_k^0)$. As in Section 6-4, we obtain this approximate interval by simply exponentiating the endpoints.

8-5 The Linear Probability Model Revisited

As we saw in Section 7-5, when the dependent variable y is a binary variable, the model must contain heteroskedasticity, unless all of the slope parameters are zero. We are now in a position to deal with this problem.

The simplest way to deal with heteroskedasticity in the linear probability model is to continue to use OLS estimation, but to also compute robust standard errors in test statistics. This ignores the fact that we actually know the form of heteroskedasticity for the LPM. Nevertheless, OLS estimation of the LPM is simple and often produces satisfactory results.