SOLUTIONS TO PROBLEMS

8.1 Parts (ii) and (iii). The homoskedasticity assumption played no role in Chapter 5 in showing that OLS is consistent. But we know that heteroskedasticity causes statistical inference based on the usual t and F statistics to be invalid, even in large samples. As heteroskedasticity is a violation of the Gauss-Markov assumptions, OLS is no longer BLUE.

8.2 \( \text{Var}(u|\text{inc}, \text{price}, \text{educ}, \text{female}) = \sigma^2 \text{inc}^2, \quad h(x) = \text{inc}^2, \) where \( h(x) \) is the heteroskedasticity function defined in equation (8.21). Therefore, \( \sqrt{h(x)} = \text{inc} \), and so the transformed equation is obtained by dividing the original equation by \( \text{inc} \):

\[
\frac{\text{beer}}{\text{inc}} = \beta_0 \left( \frac{1}{\text{inc}} \right) + \beta_1 + \beta_2 \frac{\text{price}}{\text{inc}} + \beta_3 \frac{\text{educ}}{\text{inc}} + \beta_4 \frac{\text{female}}{\text{inc}} + \left( u / \text{inc} \right).
\]

Notice that \( \beta_i \), which is the slope on \( \text{inc} \) in the original model, is now a constant in the transformed equation. This is simply a consequence of the form of the heteroskedasticity and the functional forms of the explanatory variables in the original equation.

8.3 False. The unbiasedness of WLS and OLS hinges crucially on Assumption MLR.4, and, as we know from Chapter 4, this assumption is often violated when an important variable is omitted. When MLR.4 does not hold, both WLS and OLS are biased. Without specific information on how the omitted variable is correlated with the included explanatory variables, it is not possible to determine which estimator has a small bias. It is possible that WLS would have more bias than OLS or less bias. Because we cannot know, we should not claim to use WLS in order to solve “biases” associated with OLS.
SOLUTIONS TO COMPUTER EXERCISES

C8.1 (i) Given the equation

\[ sleep = \beta_0 + \beta_1\text{totwrk} + \beta_2\text{educ} + \beta_3\text{age} + \beta_4\text{age}^2 + \beta_5\text{yngkid} + \beta_6\text{male} + u, \]

the assumption that the variance of \( u \), given all explanatory variables, depends only on gender is

\[ \text{Var}(u \mid \text{totwrk, educ, age, yngkid, male}) = \text{Var}(u \mid \text{male}) = \delta_0 + \delta_\text{male}. \]

Then the variance for women is simply \( \delta_0 \) and that for men is \( \delta_0 + \delta_1 \); the difference in variances is \( \delta_1 \).

(ii) After estimating the above equation by OLS, we regress \( \hat{u}_i^2 \) on \( \text{male}_i, i = 1, 2, \ldots, 706 \) (including, of course, an intercept). We can write the results as

\[ \hat{u}^2 = 189,359.2 - 28,849.6 \text{ male} + \text{ residual} \]

\[ (20,546.4) \quad (27,296.5) \]

\[ n = 706, \quad R^2 = .0016. \]

Because the coefficient on \( \text{male} \) is negative, the estimated variance is higher for women.

(iii) No. The \( t \) statistic on \( \text{male} \) is only about \( -1.06 \), which is not significant at even the 20% level against a two-sided alternative.
The estimated equation is

\[ \hat{\text{vote}A} = 37.66 + .252 \text{ prtystrA} + 3.793 \text{ democA} + 5.779 \log(\text{expendA}) \\
- 6.238 \log(\text{expendB}) + \hat{u} \]

\( n = 173, \ R^2 = .801, \ \overline{R}^2 = .796. \)

You can convince yourself that regressing the \( \hat{u}_i \) on all of the explanatory variables yields an \( R \)-squared of zero, although it might not be exactly zero in your computer output due to rounding error. Remember, OLS works by choosing the estimates, \( \hat{\beta}_j \), such that the residuals are uncorrelated in the sample with each independent variable (and the residuals have a zero sample average, too).

(ii) The B-P test entails regressing the \( \hat{u}_i^2 \) on the independent variables in part (i). The \( F \) statistic for joint significance (with 4 and 168 \( df \)) is about 2.33 with \( p \)-value \( \approx .058 \). Therefore, there is some evidence of heteroskedasticity, but not quite at the 5% level.

(iii) Now we regress \( \hat{u}_i^2 \) on \( \hat{\text{vote}A}_i \) and \( (\hat{\text{vote}A}_i)^2 \), where the \( \hat{\text{vote}A}_i \) are the OLS fitted values from part (i). The \( F \) test, with 2 and 170 \( df \), is about 2.79 with \( p \)-value \( \approx .065 \). This is slightly less evidence of heteroskedasticity than provided by the B-P test, but the conclusion is very similar.
(i) The estimates are given in equation (7.31). Rounded to four decimal places, the smallest fitted value is .0066 and the largest fitted value is .5577.

(ii) The estimated heteroskedasticity function for each observation \( i \) is 
\[ \hat{h}_i = arr_{86}(1 - arr_{86}), \]
which is strictly between zero and one because \( 0 < arr_{86} < 1 \) for all \( i \). The weights for WLS are \( 1/\hat{h}_i \). To show the WLS estimate of each parameter, we report the WLS results using the same equation format as for OLS:

\[
arr_{86} = 0.448 - 0.168 \text{pcnv} + 0.0054 \text{avgsen} - 0.0018 \text{tottime} - 0.025 \text{ptime86} \\
\quad (0.018) (0.019) (0.0051) (0.0033) (0.003) \]

\[ - 0.045 \text{qemp86} \]
\[ \quad (0.005) \]

\( n = 2,725, \quad R^2 = 0.0744. \)

The coefficients on the significant explanatory variables are very similar to the OLS estimates. The WLS standard errors on the slope coefficients are generally lower than the nonrobust OLS standard errors. A proper comparison would be with the robust OLS standard errors.

(iii) After WLS estimation, the \( F \) statistic for joint significance of \( \text{avgsen} \) and \( \text{tottime} \), with 2 and 2,719 \( df \), is about 0.88 with \( p \)-value \( \approx 0.41 \). They are not close to being jointly significant at the 5% level. If your econometrics package has a command for WLS and a test command for joint hypotheses, the \( F \) statistic and \( p \)-value are easy to obtain. Alternatively, you can obtain the restricted \( R \)-squared using the same weights as in part (ii) and dropping \( \text{avgsen} \) and \( \text{tottime} \) from the WLS estimation. (The unrestricted \( R \)-squared is 0.0744.)

(\( C8.7 \)) (i) The heteroskedasticity-robust standard error for \( \hat{\beta}_{\text{white}} \approx 0.129 \) is about 0.026, which is notably higher than the nonrobust standard error (about 0.020). The heteroskedasticity-robust 95% confidence interval is about 0.078 to 0.179, while the nonrobust CI is, of course, narrower, about 0.090 to 0.168. The robust CI still excludes the value zero by some margin.

(ii) There are no fitted values less than zero, but there are 213 greater than one. Unless we do something to those fitted values, we cannot directly apply WLS, as \( \hat{h}_i \) will be negative in 213 cases.