5.12 Consider the .95 level confidence region for the parameters β_1 and β_2 of the regression model (5.59). In the two-dimensional space $\mathcal{S}(\boldsymbol{x}_1, \boldsymbol{x}_2)$ generated by the two regressors, consider the set of points of the form $\beta_{10}\boldsymbol{x}_1 + \beta_{20}\boldsymbol{x}_2$, where (β_{10}, β_{20}) belongs to the confidence region. Show that this set is a circular disk with center at the OLS estimates $(\boldsymbol{x}_1\hat{\beta}_1 + \boldsymbol{x}_2\hat{\beta}_2)$. What is the radius of the disk?

5.13 Using the data in the file **earnings.data**, regress y on all three dummy variables, and compute a heteroskedasticity-consistent standard error for the coefficient of d_3. Using these results, construct a .95 asymptotic confidence interval for the mean income of individuals that belong to age group 3. Compare this interval with the ones you constructed in Exercises 5.3, 5.4, and 5.6.

5.14 Consider the linear regression model

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{u}, \quad \mathrm{E}(\boldsymbol{u}\boldsymbol{u}^\top) = \boldsymbol{\Omega},$$

where the number of observations, n, is equal to $3m$. The first three rows of the matrix $\boldsymbol{X}$ are

$$\begin{bmatrix} 1 & 4 \\ 1 & 8 \\ 1 & 15 \end{bmatrix},$$

and every subsequent group of three rows is identical to this first group. The covariance matrix $\boldsymbol{\Omega}$ is diagonal, with typical diagonal element equal to $\omega^2 x_{t2}^2$, where $\omega > 0$, and x_{t2} is the t^{th} element of the second column of $\boldsymbol{X}$.

What is the variance of $\hat{\beta}_2$, the OLS estimate of β_2? What is the probability limit, as $n \to \infty$, of the ratio of the conventional estimate of this variance, which incorrectly assumes homoskedasticity, to a heteroskedasticity-consistent estimate based on (5.39)?

5.15 Generate N simulated data sets, where N is between 1000 and 1,000,000, depending on the capacity of your computer, from each of the following two data generating processes:

$$\begin{aligned} &\text{DGP 1:} \quad y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + u_t, \quad u_t \sim \mathrm{N}(0, 1) \\ &\text{DGP 2:} \quad y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + u_t, \quad u_t \sim \mathrm{N}(0, \sigma_t^2), \;\; \sigma_t^2 = \big(\mathrm{E}(y_t)\big)^2. \end{aligned}$$

There are 50 observations, $\boldsymbol{\beta} = [1 \;\vdots\; 1 \;\vdots\; 1]$, and the data on the exogenous variables are to be found in the file **mw.data**. These data were originally used by MacKinnon and White (1985).

For each of the two DGPs and each of the N simulated data sets, construct .95 confidence intervals for β_1 and β_2 using the usual OLS covariance matrix and the HCCMEs HC_0, HC_1, HC_2, and HC_3. The OLS interval should be based on the Student's t distribution with 47 degrees of freedom, and the others should be based on the $\mathrm{N}(0, 1)$ distribution. Report the proportion of the time that each of these confidence intervals included the true values of the parameters.

On the basis of these results, which covariance matrix estimator would you recommend using in practice?