Econ 203B: Single Equation Models

Solutions for Problem Set 4

Gissele Gajate/Michael Powell
Department of Economics, UCLA
February 16th, 2006

1 Hayashi Chapter 1, Empirical Exercise

Read Marc Nerlove, "Returns to Scale in Electricity Supply" (except paragraphs of equations (6) – (9), the part of section 2 from *p.* 184 on, and Appendix A and C) before doing this exercise. For 145 electric utility companies in 1955, the file NERLOVE.ASC has data on the following:

Column 1: Total costs (call it TC) in millions of dollars
Column 2: output (Q) in billions of kilowatt hours
Column 3: price of labor (PL)
Column 4: price of fuels (PF)
Column 5: price of capital (PK).

they are from the data appendix of his article. There are 145 observations, and the observations are ordered in size, observation 1 being the smallest company and observation 145 the largest. Using the data transformation facilities of your computer software, generate for each of the 145 firms the variables required for estimation. To estimate (1.7.4), for example, you need to generate $\log(TC)$, a constant, $\log(Q)$, $\log(PL)$, $\log(PK)$, and $\log(PF)$, for each of the 145 firms.

a. (Data Question) Does Nerlove's construction of the price of capital conform to the definition of the user cost of capital? **Hint:** Read Nerlove's Appendix B.4.

Solution According to Nerlove's Appendix B.4., "First, an estimate of the current long-term rate at which the firm could borrow was obtained... This rate was in turn multiplied by the Handy-Whitman Index of Electric Utility Construction Costs for the region in which the firm had the bulk of its operations." This definition suggests that the user cost of capital is rp_I, where r is the real interest rate, and p_I is the price of capital goods. This conflicts with the definition given in Hayashi (p. 60): $(r+\delta)\,p_I$, where δ is the depreciation rate.

b. Estimate the unrestricted model (1.7.4) by OLS. Can you replicate the estimates in the text?

Solution Estimating the model given by

$$\log(TC_i) = \beta_1 + \beta_2 \log(Q_i) + \beta_3 \log(PL_i) + \beta_4 \log(PK_i) + \beta_5 \log(PF_i) + \varepsilon_i$$

We have, using MATLAB (standard errors in parenthesis):

$$\begin{aligned} \widehat{\log(TC_i)} &= \underset{(1.7744)}{-3.5265} + \underset{(0.0173)}{0.7204}\log(Q_i) + \underset{(0.2910)}{0.4363}\log(PL_i) \\ &\quad \underset{(0.3394)}{-0.2199}\log(PK_i) + \underset{(0.4265)}{0.4265}\log(PK_i) \end{aligned}$$

I will defer the discussion of the comparisons to the results given in the paper until a later section.

c. (Restricted least squares) Estimate the restricted model (1.7.6) by OLS. To do this, you need to generate a new sert of variables for each of the 145 firms. For example, the dependent variable is $\log(TC/PF)$, not $\log(TC)$. Can you replicate the estimates in the text? Can you replicate Nerlove's results? Nerlove's estimate of β_2, for example, is 0.721 with a standard error of 0.0174 (the standard error in his paper is 0.175, but it is probably a typographical error). Where in Nerlove's paper can you find this estimate? What about the other coefficients? (Warning: You will not be able to replicate

Nerlove's results precisely. One reason is that he used common rather than natural logarithms; however, this should affect only the estimated intercept term. The other reason: the data set used for his results is a corrected version of the data set published with his article.)

Solution Here, we are asked to estimate the following model:

$$\log\left(\frac{TC_i}{PF_i}\right) = \beta_1 + \beta_2 \log(Q_i) + \beta_3 \log\left(\frac{PL_i}{PF_i}\right) + \beta_4 \log\left(\frac{PK_i}{PF_i}\right) + \varepsilon_i$$

Which yields the following results:

$$\log\widehat{\left(\frac{TC_i}{PF_i}\right)} = \underset{(0.8849)}{-4.6908} + \underset{(0.0174)}{0.7206} \log(Q_i) + \underset{(0.2046)}{0.5929} \log\left(\frac{PL_i}{PF_i}\right) - \underset{(0.1907)}{0.0074} \log\left(\frac{PK_i}{PF_i}\right) + \varepsilon_i$$

Comparing this to the results given in Nerlove's paper (on page 176), we have:

$$\begin{aligned}
\hat{\beta}_2^{NERLOVE} &= 0.721;\ \ \hat{\beta}_2 = 0.7206 \\
\hat{\beta}_3^{NERLOVE} &= 0.562;\ \ \hat{\beta}_3 = 0.5929 \\
\hat{\beta}_4^{NERLOVE} &= -0.003;\ \ \hat{\beta}_4 = -0.0074
\end{aligned}$$

For the most part, the results are similar, though not equal, which is what we would expect, given the statement of the question.

As mentioned in the test, the plot of the residuals suggests a nonlinear relationship between $\log(TC)$ and $\log(Q)$. Nerlove hypothesized that estimated returns to scale varied with the level of output. Following Nerlove, divide the sample of 145 firms into five subsamples or groups, each having 29 firms. (Recall that since the data are ordered by level of output, the first 29 observations will have the smallest output levels, whereas the last 29 observations will have the largest output levels.) Consider the following three generalizations of the model (1.7.6) :

Model 1: Both the coefficients $(\beta's)$ and the error variance in (1.7.6) differ across groups.

Model 2: The coefficients are different, but the error variance is the same across groups.

Model 3: While each group has common coefficients for β_3 and β_4 (price elasticities) and common error variance, it has a different intercept term and a different β_2. Model 3 is what Nerlove called the hypothesis of neutral variations in returns to scale.

For Model 1, the coefficients and error variances specific to groups can be estimated from

$$y^{(j)} = X^{(j)}\beta^{(j)} + \varepsilon^{(j)},\ (j = 1, \ldots, 5),$$

where $y^{(j)}$ (29×1) is the vector of the values of the dependent variable for group j, $X^{(j)}$ (29×4) is the matrix of the values of the four regressors for group j, $\beta^{(j)}$ (4×1) is the coefficient vector for group j, and $\varepsilon^{(j)}$ (29×1) is the error vector. The second column of $X^{(5)}$, for example, is $\log(Q)$ for $i = 117, \ldots, 145$. Model 1 assumes conditional homoskedasticity $E\left(\varepsilon^{(j)}\varepsilon^{(j)\prime}\middle| X^{(j)}\right) = \sigma_j^2 I_{29}$ within (but not necessarily across) groups.

d. Estimate Model 1 by OLS. How well can you replicate Nerlove's reported results? On the basis of your estimates of β_2, compute the point estimates of returns to scale in each of the five groups. What is the general pattern of estimated scale economies as the level of output increases? What is the general pattern of the estimated error variance as output increases?

Solution Here, we are asked to estimate the following five models:

$$y_i^{(1)} = \beta_1^{(1)} + \beta_2 X_{2,i}^{(1)} + \beta_3 X_{3,i}^{(1)} + \beta_4 X_{4,i}^{(1)} + \varepsilon_i^{(1)}$$

$$y_i^{(2)} = \beta_1^{(2)} + \beta_2 X_{2,i}^{(2)} + \beta_3 X_{3,i}^{(2)} + \beta_4 X_{4,i}^{(2)} + \varepsilon_i^{(2)}$$

$$y_i^{(3)} = \beta_1^{(3)} + \beta_2 X_{2,i}^{(3)} + \beta_3 X_{3,i}^{(3)} + \beta_4 X_{4,i}^{(3)} + \varepsilon_i^{(3)}$$
$$y_i^{(4)} = \beta_1^{(4)} + \beta_2 X_{2,i}^{(4)} + \beta_3 X_{3,i}^{(4)} + \beta_4 X_{4,i}^{(4)} + \varepsilon_i^{(4)}$$
$$y_i^{(5)} = \beta_1^{(5)} + \beta_2 X_{2,i}^{(5)} + \beta_3 X_{3,i}^{(5)} + \beta_4 X_{4,i}^{(5)} + \varepsilon_i^{(5)}$$

Which yields the following estimates:

$$\widehat{y_i^{(1)}} = \underset{(3.1457)}{-3.3433} + \underset{(0.0845)}{0.4003} X_{2,i}^{(1)} + \underset{(0.7293)}{0.6152} X_{3,i}^{(1)} - \underset{(0.7064)}{0.0814} X_{4,i}^{(1)}$$

$$\widehat{y_i^{(2)}} = \underset{(1.4129)}{-6.4890} + \underset{(0.1163)}{0.6582} X_{2,i}^{(2)} + \underset{(0.2743)}{0.0938} X_{3,i}^{(2)} + \underset{(0.2765)}{0.3779} X_{4,i}^{(2)}$$

$$\widehat{y_i^{(3)}} = \underset{(1.6890)}{-7.3329} + \underset{(0.1980)}{0.9383} X_{2,i}^{(3)} + \underset{(0.1994)}{0.4023} X_{3,i}^{(3)} + \underset{(0.1870)}{0.2500} X_{4,i}^{(3)}$$

$$\widehat{y_i^{(4)}} = \underset{(1.1648)}{-6.5460} + \underset{(0.1075)}{0.9120} X_{2,i}^{(4)} + \underset{(0.1875)}{0.5070} X_{3,i}^{(4)} + \underset{(0.1641)}{0.0934} X_{4,i}^{(4)}$$

$$\widehat{y_i^{(5)}} = \underset{(1.0463)}{-6.7143} + \underset{(0.0650)}{1.1044} X_{2,i}^{(5)} + \underset{(0.1973)}{0.6026} X_{3,i}^{(5)} - \underset{(0.1749)}{0.2894} X_{4,i}^{(5)}$$

Comparing these results to those in Nerlove:

Obs 1-29	Nerlove			Our Estimates		
	Estimate	SE	R^2	Estimate	SE	R^2
β_2	0.398	0.079		0.400	0.084	
β_3	0.641	0.691	0.512	0.615	0.729	0.513
β_4	-0.093	0.669		-0.081	0.706	

Obs 30-58	Nerlove			Our Estimates		
	Estimate	SE	R^2	Estimate	SE	R^2
β_2	0.668	0.116		0.658	0.116	
β_3	0.105	0.275	0.635	0.094	0.274	0.633
β_4	0.364	0.227		0.378	0.277	

Obs 59-87	Nerlove			Our Estimates		
	Estimate	SE	R^2	Estimate	SE	R^2
β_2	0.931	0.198		0.938	0.198	
β_3	0.408	0.199	0.571	0.402	0.199	0.573
β_4	0.249	0.189		0.250	0.187	

Obs 88-116	Nerlove			Our Estimates		
	Estimate	SE	R^2	Estimate	SE	R^2
β_2	0.915	0.108		0.912	0.107	
β_3	0.472	0.174	0.871	0.507	0.187	0.873
β_4	0.133	0.157		0.093	0.164	

Obs 117-145	Nerlove			Our Estimates		
	Estimate	SE	R^2	Estimate	SE	R^2
β_2	1.045	0.065		1.044	0.065	
β_3	0.604	0.197	0.92	0.603	0.197	0.921
β_4	-0.295	0.175		-0.289	0.174	

Estimates of the returns to scale for each group:

Observation	$\frac{1}{\beta_2}$	SSR
1-29	2.5	8.885
30-58	1.52	1.469
59-87	1.07	0.98
88-116	1.1	0.364
117-145	0.96	0.564

It can be seen that as the level of output increase,the returns to scale decline, though not at a constant rate. The rate at which the returns to scale declines is smaller for higher levels of output. Further, it can be seen that SSR decreases significantly as the output increases, which suggests that there is less variability in the residuals as the output increases.

Model 2 assumes for Model 1 that $\sigma_j^2 = \sigma^2$ for all j. This equivariance restriction can be incorporated by stacking vectors and matrices as follows:

$$y = X\beta + \varepsilon$$

where

$$\underset{(145\times 1)}{y} = \begin{bmatrix} y^{(1)} \\ \vdots \\ y^{(5)} \end{bmatrix}, \quad \underset{(145\times 20)}{X} = \begin{bmatrix} X^{(1)} & & \\ & \ddots & \\ & & X^{(5)} \end{bmatrix}, \quad \underset{(145\times 1)}{\varepsilon} = \begin{bmatrix} \varepsilon^{(1)} \\ \vdots \\ \varepsilon^{(5)} \end{bmatrix}. \qquad (*)$$

In particular, X is now a block-diagonal matrix. The equivariance restriction can be expressed as $E(\varepsilon\varepsilon'|X) = \sigma^2 I_{145}$. There are now 20 variables derived from the original four regressors. The 145 dimensional vector corresponding to the second variable, for example, has $\log(Q_1), \ldots, \log(Q_{29})$ as the first 29 elements and zeros elsewhere. The vector corresponding to the 6th variable, which represents log output for the second group of firms, has $\log(Q_{30}), \ldots, \log(Q_{58})$ for the 30th through 58th elements and zeros elsewhere, and so on.

The stacking operation needed to form the y and X in $(*)$ can be done easily if your computer software is matrix-based. otherwise, you trick your software into accomplishing the same thing by the use of dummy variables. Define the $j-th$ dummy variable as

$$D_{ji} = \begin{cases} 1 & \text{if firm } i \text{ belongs to the } j-th \text{ group,} \\ 0 & \textit{otherwise}, \end{cases} \qquad (i = 1, \ldots, 145).$$

Then the second regressor is $D_{1i} \cdot \log(Q_i)$. The 6th variable is $D_{2i} \cdot \log(Q_i)$, and so forth.

e. Estimate Model 2 by OLS. Verify that the OLS coefficient estimates here are the same as those in (d). Also verify that

$$\sum_{j=1}^{5} SSR_j = SSR,$$

where SSR_j is the SSR from the $j-th$ group in your estimation of Model 1 in (d) and SSR is the SSR from Model 2. This agreement is not by accident, i.e., not specific to the present data set. Prove that this agreement for the coefficients and the SSR holds in general, temporarily assuming just two groups without loss of generality. **Hint:** First show that the coefficient estimate is the same between Model 1 and Model 2. Use formulas $(A.4)$, $(A.5)$, and $(A.9)$ of Appendix A.

Solution Here, we are estimating the model:

$$\begin{bmatrix} y^{(1)} \\ \vdots \\ y^{(5)} \end{bmatrix} = \begin{bmatrix} X^{(1)} & & \\ & \ddots & \\ & & X^{(5)} \end{bmatrix} \begin{bmatrix} \beta^{(1)} \\ \vdots \\ \beta^{(5)} \end{bmatrix} + \begin{bmatrix} \varepsilon^{(1)} \\ \vdots \\ \varepsilon^{(5)} \end{bmatrix}$$

From which we obtain the result (using MATLAB) that $SSR = 12.2624$. Comparing this to the results from part (d) :

$$\begin{aligned} & SSR^{(1)} + SSR^{(2)} + SSR^{(3)} + SSR^{(4)} + SSR^{(5)} \\ = \; & 8.8852 + 1.4691 + 0.9802 + 0.3636 + 0.5644 = 12.2624 \end{aligned}$$

which verifies the claim that

$$\sum_{j=1}^{5} SSR^{(j)} = SSR$$

To prove this in general, consider the regression

$$\begin{bmatrix} y^{(1)} \\ \vdots \\ y^{(5)} \end{bmatrix} = \begin{bmatrix} X^{(1)} & & \\ & \ddots & \\ & & X^{(5)} \end{bmatrix} \begin{bmatrix} \beta^{(1)} \\ \vdots \\ \beta^{(5)} \end{bmatrix} + \begin{bmatrix} \varepsilon^{(1)} \\ \vdots \\ \varepsilon^{(5)} \end{bmatrix}$$

From which we obtain

$$\begin{aligned} \hat{\beta} &= \begin{bmatrix} \hat{\beta}^{(1)} \\ \vdots \\ \hat{\beta}^{(5)} \end{bmatrix} = \left(X'X\right)^{-1} X'y \\ &= \begin{bmatrix} \left(X^{(1)\prime} X^{(1)}\right)^{-1} & & \\ & \ddots & \\ & & \left(X^{(5)\prime} X^{(5)}\right)^{-1} \end{bmatrix} \begin{bmatrix} X^{(1)\prime} y^{(1)} \\ \vdots \\ X^{(5)\prime} y^{(5)} \end{bmatrix} \\ &= \begin{bmatrix} \left(X^{(1)\prime} X^{(1)}\right)^{-1} X^{(1)\prime} y^{(1)} \\ \vdots \\ \left(X^{(5)\prime} X^{(5)}\right)^{-1} X^{(5)\prime} y^{(5)} \end{bmatrix} \end{aligned}$$

The residuals would therefore be:

$$\hat{\varepsilon} = \begin{bmatrix} \hat{\varepsilon}^{(1)} \\ \vdots \\ \hat{\varepsilon}^{(2)} \end{bmatrix} = \begin{bmatrix} y^{(1)} - X^{(1)} \left(X^{(1)\prime} X^{(1)}\right)^{-1} X^{(1)\prime} y^{(1)} \\ \vdots \\ y^{(5)} - X^{(5)} \left(X^{(5)\prime} X^{(5)}\right)^{-1} X^{(5)\prime} y^{(5)} \end{bmatrix}$$

But note that these estimates are exactly those we would obtain by running the individual regressions and therefore,

$$SSR = \hat{\varepsilon}'\hat{\varepsilon} = \begin{bmatrix} \hat{\varepsilon}^{(1)\prime} & \cdots & \hat{\varepsilon}^{(5)\prime} \end{bmatrix} \begin{bmatrix} \hat{\varepsilon}^{(1)} \\ \vdots \\ \hat{\varepsilon}^{(5)} \end{bmatrix} = \sum_{j=1}^{5} \hat{\varepsilon}^{(j)\prime} \hat{\varepsilon}^{(j)} = \sum_{j=1}^{5} SSR^{(j)}$$

Which is the desired result.

f. (Chow test) Model 2 is more general than Model (1.7.6) because the coefficients can differ across groups. Test the null hypothesis that the coefficients are the same across groups. How many equations (restrictions) are in the null hypothesis? This test is sometimes called the **Chow test for structural change**. Calculate the p-value of the $F-$ratio. **Hint:** This is a linear hypothesis about the coefficients of Model 2. So take Model 2 to be the maintained hypothesis and (1.7.6) to be the restricted model. Use the formula (1.4.11) for the F-ratio.

Solution Here, the null hypothesis is $H_0 : \beta^{(1)} = \beta^{(2)} = \beta^{(3)} = \beta^{(4)} = \beta^{(5)}$, which amounts to sixteen equalities since $\beta^{(j)}$ is 4×1 for each j. (Hence, $p = 16$) The restricted model from part (c) yields $SSR_R = 21.640$, and the unrestricted model from part (e) yields $SSR_U = 12.262$. Using the appropriate F statistic, with $n = 145$ and $k = 20$, we have:

$$F_0 = \frac{(SSR_R - SSR_U)/p}{SSR_U/(n-k)} = \frac{(21.640 - 12.262)/16}{12.262/125} = 5.975$$

The corresponding critical value is $c^*_{0.05,F(16,125)} = 1.73$. Since $F_0 > c^*_{0.05,F(16,125)}$, we reject the null hypothesis in favor of structural change at the 5% significance level.

The restriction in Model 3 that the price elasticities are the same across firm groups can be imposed on Model 2 by applying the dummy variable transformation only to the constant and log output. Thus, there are 12 ($= 2 \times 5 + 2$) variables in X. Now X looks like

$$X = \begin{bmatrix} 1 & \log(Q_1) & 0 & 0 & \log(PL_1/PF_1) & \log(PK_1/PF_1) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \log(Q_{29}) & 0 & 0 & \log(PL_{29}/PF_{29}) & \log(PK_{29}/PF_{29}) \\ & & \ddots & & \vdots & \vdots \\ 0 & 0 & 1 & \log(Q_{117}) & \log(PL_{117}/PF_{117}) & \log(PK_{117}/PF_{117}) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & \log(Q_{145}) & \log(PL_{145}/PF_{145}) & \log(PK_{145}/PF_{145}) \end{bmatrix} \quad (**)$$

g. Estimate Model 3. The model is a special case of Model 2, with the hypothesis that the two price elasticities are the same across the five groups. Test the hypothesis at a significance level of 5 percent, assuming normality. (Note: Nerlove's F-ratio on p. 183 is wrong.)

Solution The hypothesis we are testing here is

$$H_0 : \beta_3^{(1)} = \beta_3^{(2)} = \beta_3^{(3)} = \beta_3^{(4)} = \beta_3^{(5)},\ \beta_4^{(1)} = \beta_4^{(2)} = \beta_4^{(3)} = \beta_4^{(4)} = \beta_4^{(5)}$$

Which amounts to eight equalities. (Giving us $p = 8$). The unrestricted model is the one we estimated in part (e) which has $SSR_U = 12.262$. Estimating the restricted model (using MATLAB) gives us $SSR_R = 12.577$. Once again, we have that $n = 145$ and $k = 20$. Putting this together, we get the following test statistic:

$$F_0 = \frac{(SSR_R - SSR_U)/p}{SSR_U /\!/ (n-k)} = \frac{(12.577 - 12.262)/8}{12.262/(125)} = 0.4012$$

The critical value is $c^*_{0.05,F(8,125)} = 2.01$ and, since $F_0 \le c^*_{0.05,F(8,125)}$, we fail to reject the null that the two price elasticities are the same across the five groups.

As has become clear from the plot of residuals in Figure 1.7, the conditional second moment $E\left(\varepsilon_i^2 \middle| X\right)$ is likely to depend on log output, which is a violation of the conditional homoskedasticity assumption. this time we do not attempt to test conditional homoskedasticity, because to do so requires large sample theory and is postponed until next chapter. Instead, we pretend to know the form of the function linking the conditional second moment to log output. The function, specified below, implies that the conditional second moment varies continuously with output, contrary to the three models we have considered above. Also contrary to these models, we assume that the degree of returns to scale varies continuously with output by including the square of log output. Model 4 is:

Model 4:

$$\begin{aligned} \log\left(\frac{TC_i}{p_{i3}}\right) &= \beta_1 + \beta_2 \log(Q_i) + \beta_3 \left[\log(Q_i)\right]^2 \\ &\quad + \beta_4 \log\left(\frac{p_{i1}}{p_{i3}}\right) + \beta_5 \log\left(\frac{p_{i2}}{p_{i3}}\right) + \varepsilon_i \end{aligned}$$

$$E\left(\varepsilon_i^2 \middle| X\right) = \sigma^2 \cdot \left(0.0565 + \frac{2.1377}{Q_i}\right) \quad (i = 1, 2, \ldots, 145)$$

for some unknown σ^2.

h. Estimate Model 4 by weighted least squares on the whole sample of 145 firms. (Be careful about the treatment of the intercept; in the equation after weighting, none of the regressors is a constant.) Plot the residuals. Is there still evidence for conditional homoskedasticity or further nonlinearities?

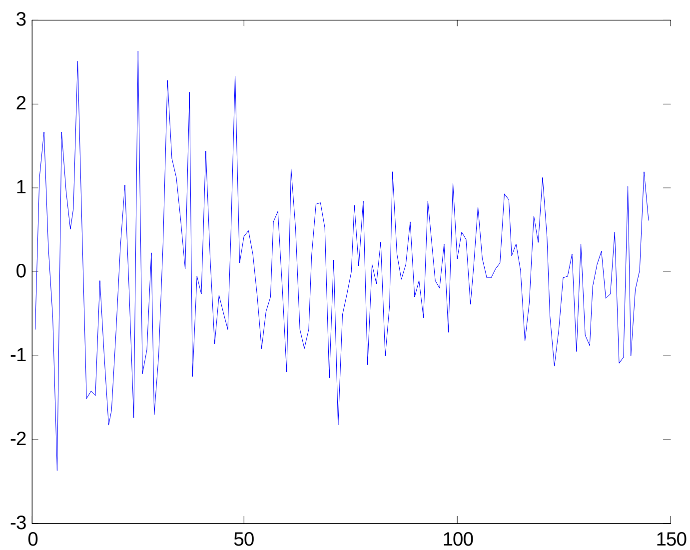
Solution As assumed, we know that $E\left[\varepsilon_i^2 \middle| X\right] = \sigma^2 \cdot \left(0.0565 + \frac{2.1377}{Q_i}\right)$. Let $v_i(Q_i) = 0.0565 + \frac{2.1377}{Q_i}$ so that $E\left[\varepsilon^2 \middle| X\right] = \sigma^2 \cdot \begin{bmatrix} v_1(Q_1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & v_{145}(Q_{145}) \end{bmatrix}$. Performing weighted least squares by dividing each variable by the square root of $v_i(Q_i)$ and estimating using OLS:

$$\begin{aligned} \frac{\log\left(\frac{TC_i}{p_{i3}}\right)}{\sqrt{v_i(Q_i)}} &= \beta_1 \frac{1}{\sqrt{v_i(Q_i)}} + \beta_2 \frac{\log(Q_i)}{\sqrt{v_i(Q_i)}} + \beta_3 \frac{[\log(Q_i)]^2}{\sqrt{v_i(Q_i)}} \\ &\quad + \beta_4 \frac{\log\left(\frac{p_{i1}}{p_{i3}}\right)}{\sqrt{v_i(Q_i)}} + \beta_5 \frac{\log\left(\frac{p_{i2}}{p_{i3}}\right)}{\sqrt{v_i(Q_i)}} + \varepsilon_i \end{aligned}$$

We get the following results (using MATLAB):

$$\begin{aligned} \widehat{\frac{\log\left(\frac{TC_i}{p_{i3}}\right)}{\sqrt{v_i(Q_i)}}} &= \underset{(0.5791)}{-4.1256} \frac{1}{\sqrt{v_i(Q_i)}} + \underset{(0.0819)}{0.2177} \frac{\log(Q_i)}{\sqrt{v_i(Q_i)}} + \underset{(0.0062)}{0.0457} \frac{[\log(Q_i)]^2}{\sqrt{v_i(Q_i)}} \\ &\quad + \underset{(0.1214)}{0.4662} \frac{\log\left(\frac{p_{i1}}{p_{i3}}\right)}{\sqrt{v_i(Q_i)}} + \underset{(0.1126)}{0.1335} \frac{\log\left(\frac{p_{i2}}{p_{i3}}\right)}{\sqrt{v_i(Q_i)}} + \varepsilon_i \end{aligned}$$

Plotting the residuals,



The plot shows that weighted least squares has taken care of most of the nonlinearities; however, there appears to still be some heteroskedasticity, since there is more variation in the residuals for lower levels of output.

2 Additional Problems

1. This problem asks you to perform a Monte Carlo exercise to study the finite sample properties of the OLS and GLS estimators, the estimators of their standard errors and of associated $t-$statistics.

a. Generate 1000 samples of size n for (y_i, x_i) from the (homoskedastic) model

$$y_i = x_i\beta + \varepsilon_i$$

where $x_i \sim N(0,1)$, $\varepsilon_i \sim N\left(0,\sigma^2\right)$ and $\beta = 1$. Calculate the OLS estimator, its standard errors, and its associated $t-$statistics for testing the null hypothesis $H_0 : \beta = 1$ both under the assumption of homoskedasticity and under the assumption of (unknown form) of heteroskedasticity. Then examine the empirical size of your two $t-$tests at the 5% and 10% nominal levels by calculating the percentage of times the $t-$statistics reject the null using the appropriate 5 and 10% critical values. Do this for $n = 100,\ 400,\ 1600$, and for $\sigma^2 = 0.5,\ 1.0$ and 2.0.

Solution We estimate

$$\begin{aligned}
\widehat{\beta}_{OLS} &= (X'X)^{-1}X'Y \\
\widehat{\varepsilon}_{OLS} &= Y - X\widehat{\beta}_{OLS} \\
t_{OLS} &= \frac{\widehat{\beta}_{OLS} - 1}{\sqrt{\widehat{\sigma}^2_{OLS}(X'X)^{-1}}} \\
where\ \widehat{\sigma}^2_{OLS} &= \frac{\widehat{\varepsilon}'_{OLS}\widehat{\varepsilon}_{OLS}}{n-1}
\end{aligned}$$

and calculate the percentage of times we reject the null hypothesis that $\beta = 1$ under the assumption of homoskedasticity, while we know that the true generated data are homoskedastic. All the tables of results are attached.

Under the assumption of heteroskedasticity, we must calculate robust standard errors. To do this, we use the asymptotic variance of the OLS estimator, which is

$$\frac{1}{n}AV(\widehat{\beta}_{OLS}) = (\sum_{i=1}^{n} X_i'X_i)^{-1}(\sum_{i=1}^{n} X_i'X_i\varepsilon_i^2)(\sum_{i=1}^{n} X_i'X_i)^{-1}$$

and from which we can calculate the standard error of the estimator,

$$SE(\widehat{\beta}_{OLS}) = \left((\sum_{i=1}^{n} X_i'X_i)^{-1}(\sum_{i=1}^{n} X_i'X_i\varepsilon_i^2)(\sum_{i=1}^{n} X_i'X_i)^{-1}\right)^{\frac{1}{2}}$$

Our t-statistic is then

$$t_{OLS,hetero} = \frac{\widehat{\beta}_{OLS} - 1}{\sqrt{SE(\widehat{\beta}_{OLS})(X'X)^{-1}}}$$

Please see the attached tables for the data and graphs.

b. Repeat the exercise above for $\varepsilon_i \sim N\left(0, \sigma_i^2\right)$ where $\sigma_i^2 = x_i^2/c$ for $c = 1, 2$ and 4.

Solution We make the same calculations in this part except that the true generated data are heteroskedastic.

c. For the two sets of Monte Carlo experiments of parts (a) and (b) compute the (infeasible) and the feasible GLS estimators under the assumption that $\sigma_i^2 = x_i^2/c$. Calculate the mean and median bias, root mean squared error and median absolute error and graph the empirical distribution of the GLS and FGLS estimators.

Solution In this section, we calculate the infeasible GLS estimator using the true variance to weight the data and the feasible GLS estimator using the estimated variance to weight the data. The new regression looks like this, with the estimates replacing the true variance for the feasible estimator

$$\frac{Y_i}{\sqrt{\sigma_i^2}} = \frac{X_i}{\sqrt{\sigma_i^2}}\beta + \frac{\varepsilon_i}{\sqrt{\sigma_i^2}}$$

When we assume a functional form to the variance, we have that

$$\frac{Y_i\sqrt{c}}{\sqrt{X_i^2}} = \frac{X_i\sqrt{c}}{\sqrt{X_i^2}}\beta + \frac{\varepsilon_i\sqrt{c}}{\sqrt{X_i^2}}$$

while when we do not have the functional form, we must estimate the variance, using

$$\sqrt{\widehat{\sigma}_i^2} = \left(\frac{\Sigma(Y_i - X_i\widehat{\beta}_{GLS})^2}{n-1}\right)^{\frac{1}{2}}$$

leaving us with

$$\frac{Y_i}{\sqrt{\widehat{\sigma}_i^2}} = \frac{X_i}{\sqrt{\widehat{\sigma}_i^2}}\beta + \frac{\varepsilon_i}{\sqrt{\widehat{\sigma}_i^2}}$$

The calculations follow in the tables.

d. For the two sets of Monte Carlo experiments of parts (a) and (b) compute the standard errors and the associated $t-$statistics for testing the null hypothesis $H_0 : \beta = 1$ for both the infeasible and the feasible GLS estimators. Then examine the empirical size of the $t-$tests at the 5% and 10% nominal levels by calculating the percentage of times the $t-$statistics reject the null using the appropriate 5 and 10% critical values.

Comment on your findings, especially on the relative performance of OLS and GLS (and FGLS) under both homoskedasticity and heteroskedasticity and their relative efficiency.

Solution The results follow in the tables. We can see from the tables that in the case of homoskedastic data, the OLS estimator rejects fewer of the true hypotheses than do the GLS and FGLS estimators, while in the case of heteroskedastic data, GLS and FGLS estimators reject fewer of the true hypotheses than does the OLS estimator. When correcting for spurious heteroskedasticity using robust standard errors, the OLS estimator continues to reject fewer; however, when the heteroskedasticity really exists in the data, the corrected OLS is not as efficient as the GLS and FGLS estimators.

2. Two samples of 50 observations each produce the following moment matrices. (In each case, the matrix X contains the unit vector and one variable.)

$$\begin{array}{lcll} X'X & : & \left[\begin{array}{cc} 50 & 300 \\ 300 & 2100 \end{array}\right] & \left[\begin{array}{cc} 50 & 300 \\ 300 & 2100 \end{array}\right] \\ X'Y & : & \left[\begin{array}{c} 300 \\ 2000 \end{array}\right] & \left[\begin{array}{c} 300 \\ 2200 \end{array}\right] \\ Y'Y & : & 2100 & 2800 \end{array}$$

a. Compute the least squares regression coefficients and the residual variances $\hat{\sigma}^2$ for each data set. Compute the R^2 for each regression.

Solution Here, we want to compute $\hat{\beta}_{OLS} = (X'X)^{-1} X'Y$, $\hat{\sigma}^2_{OLS} = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n-k}$, and $R^2_{UC} = 1 - \frac{\hat{\varepsilon}'\hat{\varepsilon}}{Y'Y}$ for each sample. Proceeding with the first sample, the calculation of $\hat{\beta}_{OLS}$ is rather straightforward:

$$\hat{\beta}_{OLS} = (X'X)^{-1} X'Y = \left[\begin{array}{cc} 50 & 300 \\ 300 & 2100 \end{array}\right]^{-1} \left[\begin{array}{c} 300 \\ 2000 \end{array}\right] = \left[\begin{array}{c} 2 \\ \frac{2}{3} \end{array}\right]$$

Next, notice that

$$\begin{aligned} \hat{\varepsilon}'\hat{\varepsilon} &= Y'M_X Y = Y'\left(I - X(X'X)^{-1} X'\right) Y \\ &= Y'Y - Y'X(X'X)^{-1} X'Y \\ &= Y'Y - (X'Y)'(X'X)^{-1} X'Y \end{aligned}$$

Plugging in the given matrices, we have:

$$\hat{\varepsilon}'\hat{\varepsilon} = 2100 - \left[\begin{array}{c} 300 \\ 2000 \end{array}\right]' \left[\begin{array}{cc} 50 & 300 \\ 300 & 2100 \end{array}\right]^{-1} \left[\begin{array}{c} 300 \\ 2000 \end{array}\right] = \frac{500}{3}$$

And therefore,

$$\hat{\sigma}^2_{OLS} = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n-k} = \frac{\frac{500}{3}}{50-2} = \frac{125}{36} = 3.4722$$

Finally, for this sample, we have:

$$R^2_{UC} = 1 - \frac{\hat{\varepsilon}'\hat{\varepsilon}}{Y'Y} = 1 - \frac{\frac{500}{3}}{2100} = \frac{58}{63} = 0.9206$$

Proceeding similarly for the second sample, we have:

$$\hat{\beta}_{OLS} = (X'X)^{-1} X'Y = \left[\begin{array}{cc} 50 & 300 \\ 300 & 2100 \end{array}\right]^{-1} \left[\begin{array}{c} 300 \\ 2200 \end{array}\right] = \left[\begin{array}{c} -2 \\ \frac{4}{3} \end{array}\right]$$

And,

$$\begin{aligned} \hat{\varepsilon}'\hat{\varepsilon} &= Y'Y - (X'Y)'(X'X)^{-1}(X'Y) \\ &= 2800 - \left[\begin{array}{c} 300 \\ 2200 \end{array}\right]' \left[\begin{array}{cc} 50 & 300 \\ 300 & 2100 \end{array}\right]^{-1} \left[\begin{array}{c} 300 \\ 2200 \end{array}\right] \\ &= \frac{1400}{3} \end{aligned}$$

Which leads us to:

$$\hat{\sigma}^2_{OLS} = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n-k} = \frac{\frac{1400}{3}}{50-2} = \frac{175}{18} = 9.7222$$

And

$$R_{UC}^2 = 1 - \frac{\hat{\varepsilon}'\hat{\varepsilon}}{Y'Y} = 1 - \frac{\frac{1400}{3}}{2800} = \frac{5}{6} = 0.8333$$

b. Compute the OLS estimate of the coefficient vector assuming that the coefficients and disturbance variance are the same in the two regressions. Also compute the estimate of the asymptotic covariance matrix of the estimate.

Solution Here, we are asked to pool together the data from the two 50-observation regressions. That is, let

$$Z = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}, \; W = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}$$

Where the subscript 1 corresponds to the data that generated the first set of results and 2 corresponds to the data that generated the second set of results. From this definition, we can construct the following matrices which will be helpful in answering this question:

$$\begin{aligned} Z'Z &= \begin{bmatrix} X_1' & X_2' \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = X_1'X_1 + X_2'X_2 \\ &= \begin{bmatrix} 50 & 300 \\ 300 & 2100 \end{bmatrix} + \begin{bmatrix} 50 & 300 \\ 300 & 2100 \end{bmatrix} \\ &= \begin{bmatrix} 100 & 600 \\ 600 & 4200 \end{bmatrix} \end{aligned}$$

$$\begin{aligned} Z'W &= \begin{bmatrix} X_1' & X_2' \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = X_1'Y_1 + X_2'Y_2 \\ &= \begin{bmatrix} 300 \\ 2000 \end{bmatrix} + \begin{bmatrix} 300 \\ 2200 \end{bmatrix} \\ &= \begin{bmatrix} 600 \\ 4200 \end{bmatrix} \end{aligned}$$

$$\begin{aligned} W'W &= \begin{bmatrix} Y_1' & Y_2' \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = Y_1'Y_1 + Y_2'Y_2 \\ &= 2100 + 2800 = 4900 \end{aligned}$$

It is then straightforward to compute $\hat{\beta}_{OLS}$:

$$\begin{aligned} \hat{\beta}_{OLS} &= (Z'Z)^{-1} Z'W \\ &= \begin{bmatrix} 100 & 600 \\ 600 & 4200 \end{bmatrix}^{-1} \begin{bmatrix} 600 \\ 4200 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \end{aligned}$$

What is the estimated asymptotic variance of $\hat{\beta}_{OLS}$? This requires some derivation, which is dealt with in some detail in questions 3 and 4. Without going into much detail here, recall the following train of thought:

$$\hat{\beta}_{OLS} = \beta + \left(\frac{1}{n}\sum_{i=1}^{n} Z_i'Z_i\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} Z_i'\varepsilon_i\right)$$

Or

$$\begin{aligned} \hat{\beta}_{OLS} - \beta &= \left(\frac{1}{n}\sum_{i=1}^{n} Z_i'Z_i\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} Z_i'\varepsilon_i\right) \\ \sqrt{n}\left(\hat{\beta}_{OLS} - \beta\right) &= \underbrace{\left(\frac{1}{n}\sum_{i=1}^{n} Z_i'Z_i\right)^{-1}}_{\xrightarrow{p}\left(E\left[Z_i'Z_i\right]\right)^{-1}} \underbrace{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} Z_i'\varepsilon_i\right)}_{\xrightarrow{d} N\left(0, Var\left(Z_i'\varepsilon_i\right)\right)} \end{aligned}$$

Where

$$\begin{aligned} Var\left(Z_i'\varepsilon_i\right) &= E\left[Var\left(\left.Z_i'\varepsilon_i\right| Z_i\right)\right] + Var\left(E\left[\left.Z_i'\varepsilon_i\right| Z_i\right]\right) \\ &= E\left[Z_i' Var\left(\left.\varepsilon_i\right| Z_i\right) Z_i\right] + Var\left(Z_i \underbrace{E\left[\left.\varepsilon_i\right| Z_i\right]}_{=0}\right) \\ &= \sigma^2 E\left[Z_i' Z_i\right] \end{aligned}$$

Putting this together (with the aid of Slutsky's theorem):

$$\sqrt{n}\left(\hat{\beta}_{OLS} - \beta\right) \xrightarrow{d} N\left(0, AV\left(\hat{\beta}_{OLS}\right)\right)$$

Where

$$\begin{aligned} AV\left(\hat{\beta}_{OLS}\right) &= \left(E\left[Z_i'Z_i\right]\right)^{-1}\left(\sigma^2 E\left[Z_i'Z_i\right]\right)\left(E\left[Z_i'Z_i\right]\right)^{-1} \\ &= \sigma^2\left(E\left[Z_i'Z_i\right]\right)^{-1} \end{aligned}$$

This gives us:

$$\hat{\beta}_{OLS} \overset{A}{\sim} N\left(\beta, \frac{1}{n} AV\left(\hat{\beta}_{OLS}\right)\right)$$

Clearly, since this variance is a function of both σ^2 and $E\left[Z_i'Z_i\right]$, neither of which are known, we must use a consistent estimator. (For this simple case, replace σ^2 with $\hat{\sigma}^2$ and $E\left[Z_i'Z_i\right]$ with $\frac{1}{n}\sum_{i=1}^{n} Z_i'Z_i$)

$$\begin{aligned} \widehat{Var\left(\hat{\beta}_{OLS}\right)} &= \frac{1}{n}\hat{\sigma}^2\left(\frac{1}{n}\sum_{i=1}^{n} Z_i'Z_i\right)^{-1} \\ &= \hat{\sigma}^2\left(\sum_{i=1}^{n} Z_i'Z_i\right)^{-1} = \hat{\sigma}^2\left(Z'Z\right)^{-1} \end{aligned}$$

We need only calculate $\hat{\sigma}^2$ and $\left(Z'Z\right)^{-1}$ and then we are done:

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{n-k}\hat{\varepsilon}'\hat{\varepsilon} = \frac{1}{n-k} W'\left(I - Z\left(Z'Z\right)^{-1}Z'\right)W \\ &= \frac{1}{n-k}\left(W'W - \left(Z'W\right)'\left(Z'Z\right)^{-1}\left(Z'W\right)\right) \\ &= \frac{1}{100-2}\left(4900 - \begin{bmatrix} 600 & 4200 \end{bmatrix}\begin{bmatrix} 100 & 600 \\ 600 & 4200 \end{bmatrix}^{-1}\begin{bmatrix} 600 \\ 4200 \end{bmatrix}\right) \\ &= \frac{1}{98}\left(4900 - 4200\right) = \frac{700}{98} = \frac{50}{7} \end{aligned}$$

Thus,

$$\begin{aligned} \widehat{Var\left(\hat{\beta}_{OLS}\right)} &= \hat{\sigma}^2\left(Z'Z\right)^{-1} \\ &= \frac{50}{7}\begin{bmatrix} \frac{7}{100} & -\frac{1}{100} \\ -\frac{1}{100} & \frac{1}{600} \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{2} & -\frac{1}{14} \\ -\frac{1}{14} & \frac{1}{84} \end{bmatrix} \end{aligned}$$

c. Test the hypothesis that the variances in the two regressions are the same without assuming that the coefficients are the same in the two regressions.

Solution Let σ_1^2 and σ_2^2 denote the true variances of the first and second regressions respectively. Here, we are asked to test $H_0 : \sigma_1^2 = \sigma_2^2$. This is equivalent to: $H_0 : \frac{\sigma_2^2}{\sigma_1^2} = 1$. For simplicity, assume that $H_1 : \frac{\sigma_2^2}{\sigma_1^2} > 1$. (The motivation for this particular alternative hypothesis is that, in part (a), we saw that $\hat{\sigma}_2^2 > \hat{\sigma}_1^2$, which gives some indication that it might be the case that $\sigma_2^2 > \sigma_1^2$.) Recall the following facts:

$$\begin{aligned} \frac{\hat{\varepsilon}_1'\hat{\varepsilon}_1}{\sigma_1^2} &\sim \chi^2(n_1 - k) \\ \frac{\hat{\varepsilon}_2'\hat{\varepsilon}_2}{\sigma_2^2} &\sim \chi^2(n_2 - k) \end{aligned}$$

Thus, we would expect that

$$\frac{\left(\frac{\hat{\varepsilon}_2'\hat{\varepsilon}_2}{\sigma_2^2}\right)/(n_2 - k)}{\left(\frac{\hat{\varepsilon}_1'\hat{\varepsilon}_1}{\sigma_1^2}\right)/(n_1 - k)} \sim F(n_2 - k, n_1 - k)$$

And, under the null, where $\frac{\sigma_2^2}{\sigma_1^2} = 1$ (and recognizing that for this particular question, $n_1 = n_2 = 50$):

$$F_0 = \frac{\hat{\varepsilon}_2'\hat{\varepsilon}_2}{\hat{\varepsilon}_1'\hat{\varepsilon}_1} \sim F(48, 48)$$

Therefore, if $F_0 > c^*_{0.05, F(48,48)} = 1.62$, we would reject the null that the variances are the same. Plugging in the results from part (a), we have:

$$F_0 = \frac{9.7222}{3.4722} = 2.8$$

And we therefore reject the null in favor of the alternative hypothesis that the second sample has a higher variance than the first.

d. Compute the two-step FGLS estimator of the coefficients in the regressions, assuming that the constant and slope are the same in both regressions. Compute the estimate of the covariance matrix and compare it with the result of part b.

Solution Here, the two-step FGLS process involves doing the following:

1. Use your original OLS estimate $\hat{\beta}_{OLS}$ for the pooled model to calculate the estimated variance-covariance matrix $\hat{\Omega}$ for the regression.

2. Use $\hat{\Omega}$ to compute $\hat{\beta}_{FGLS} = \left(Z'\hat{\Omega}^{-1}Z\right)^{-1} Z'\hat{\Omega}^{-1}W$.

Using $\hat{\beta}_{OLS} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ as calculated in part (b) to calculate $\hat{\Omega} = \begin{bmatrix} \hat{\sigma}_1^2 \cdot I_{50} & 0 \\ 0 & \hat{\sigma}_2^2 \cdot I_{50} \end{bmatrix}$, we have:

$$\begin{aligned} \hat{\sigma}_1^2 &= \frac{\hat{\varepsilon}_1'\hat{\varepsilon}_1}{n-k} = \frac{\left(Y_1 - X_1 \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right)' \left(Y_1 - X_1 \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right)}{n-k} \\ &= \frac{Y_1'Y_1 - \begin{bmatrix} 0 & 1 \end{bmatrix} X_1'Y_1 - \left(\begin{bmatrix} 0 & 1 \end{bmatrix} X_1'Y_1\right)' + \begin{bmatrix} 0 & 1 \end{bmatrix} X_1'X_1 \begin{bmatrix} 0 \\ 1 \end{bmatrix}}{n-k} \\ &= \frac{2100 - 2\begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} 300 \\ 2000 \end{bmatrix} + \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} 50 & 300 \\ 300 & 2100 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix}}{48} \\ &= \frac{25}{6} \end{aligned}$$

And

$$\begin{aligned}
\hat{\sigma}_2^2 &= \frac{\hat{\varepsilon}_2'\hat{\varepsilon}_2}{n-k} = \frac{\left(Y_2 - X_2\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right)'\left(Y_2 - X_2\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right)}{n-k} \\
&= \frac{Y_2'Y_2 - \begin{bmatrix} 0 & 1 \end{bmatrix} X_2'Y_2 - \left(\begin{bmatrix} 0 & 1 \end{bmatrix} X_2'Y_2\right)' + \begin{bmatrix} 0 & 1 \end{bmatrix} X_2'X_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix}}{n-k} \\
&= \frac{2800 - 2\begin{bmatrix} 0 & 1 \end{bmatrix}\begin{bmatrix} 300 \\ 2200 \end{bmatrix} + \begin{bmatrix} 0 & 1 \end{bmatrix}\begin{bmatrix} 50 & 300 \\ 300 & 2100 \end{bmatrix}\begin{bmatrix} 0 \\ 1 \end{bmatrix}}{48} \\
&= \frac{125}{12}
\end{aligned}$$

Therefore, our estimated variance-covariance matrix for the regression is:

$$\hat{\Omega} = \begin{bmatrix} \frac{25}{6} \cdot I_{50} & 0 \\ 0 & \frac{125}{12} \cdot I_{50} \end{bmatrix} = \begin{bmatrix} \hat{\sigma}_1^2 \cdot I_{50} & 0 \\ 0 & \hat{\sigma}_2^2 \cdot I_{50} \end{bmatrix}$$

Though the question does not explicitly ask us to solve for the FGLS estimates, I have done so (for fun):

$$\begin{aligned}
\hat{\beta}_{FGLS} &= \left(Z'\hat{\Omega}^{-1}Z\right)^{-1} Z'\hat{\Omega}^{-1}W \\
&= \left(\begin{bmatrix} X_1' & X_2' \end{bmatrix}\begin{bmatrix} \left(\hat{\sigma}_1^2\right)^{-1} \cdot I_{50} & 0 \\ 0 & \left(\hat{\sigma}_2^2\right)^{-1} \cdot I_{50} \end{bmatrix}\begin{bmatrix} X_1 \\ X_2 \end{bmatrix}\right)^{-1} \\
&\quad \cdot \begin{bmatrix} X_1' & X_2' \end{bmatrix}\begin{bmatrix} \left(\hat{\sigma}_1^2\right)^{-1} \cdot I_{50} & 0 \\ 0 & \left(\hat{\sigma}_2^2\right)^{-1} \cdot I_{50} \end{bmatrix}\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \\
&= \left(\frac{1}{\hat{\sigma}_1^2}X_1'X_1 + \frac{1}{\hat{\sigma}_2^2}X_2'X_2\right)^{-1}\left(\frac{1}{\hat{\sigma}_1^2}X_1'Y_1 + \frac{1}{\hat{\sigma}_2^2}X_2'Y_2\right) \\
&= \left(\frac{6}{25}\begin{bmatrix} 50 & 300 \\ 300 & 2100 \end{bmatrix} + \frac{12}{125}\begin{bmatrix} 50 & 300 \\ 300 & 2100 \end{bmatrix}\right)^{-1} \\
&\quad \cdot \left(\frac{6}{25}\begin{bmatrix} 300 \\ 2000 \end{bmatrix} + \frac{12}{125}\begin{bmatrix} 300 \\ 2200 \end{bmatrix}\right) \\
&\quad = \begin{bmatrix} \frac{5}{12} & -\frac{5}{84} \\ -\frac{5}{84} & \frac{5}{504} \end{bmatrix}\begin{bmatrix} \frac{504}{5} \\ \frac{3456}{5} \end{bmatrix} = \begin{bmatrix} \frac{6}{7} \\ \frac{6}{7} \end{bmatrix}
\end{aligned}$$

Our variance-covariance matrix for $\hat{\beta}_{FGLS}$ is therefore:

$$\begin{aligned}
Var\left(\hat{\beta}_{FGLS}\right) &= Var\left(\left(Z'\Omega^{-1}Z\right)^{-1} Z'\Omega^{-1}W\right) \\
&= \left(Z'\Omega^{-1}Z\right)^{-1} Z'\Omega^{-1}\underbrace{Var\left(W\right)}_{\Omega}\Omega^{-1}Z\left(Z'\Omega^{-1}Z\right)^{-1} \\
&= \left(Z'\Omega^{-1}Z\right)^{-1}
\end{aligned}$$

Which can be estimated with:

$$\begin{aligned}
\widehat{Var\left(\hat{\beta}_{FGLS}\right)} &= \left(Z'\hat{\Omega}^{-1}Z\right)^{-1} \\
&= \left(\begin{bmatrix} X_1' & X_2' \end{bmatrix} \begin{bmatrix} \left(\hat{\sigma}_1^2\right)^{-1} \cdot I_{50} & 0 \\ 0 & \left(\hat{\sigma}_2^2\right)^{-1} \cdot I_{50} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}\right)^{-1} \\
&= \left(\frac{6}{25}\begin{bmatrix} 50 & 300 \\ 300 & 2100 \end{bmatrix} + \frac{12}{125}\begin{bmatrix} 50 & 300 \\ 300 & 2100 \end{bmatrix}\right)^{-1} \\
&= \begin{bmatrix} \frac{5}{12} & -\frac{5}{84} \\ -\frac{5}{84} & \frac{5}{504} \end{bmatrix}
\end{aligned}$$

What, then, is the relationship between $\widehat{Var\left(\hat{\beta}_{FGLS}\right)}$ and $\widehat{Var\left(\hat{\beta}_{OLS}\right)}$, which was calculated in part (b)?

Taking the difference:

$$\begin{aligned}
\Delta &\equiv \widehat{Var\left(\hat{\beta}_{OLS}\right)} - \widehat{Var\left(\hat{\beta}_{FGLS}\right)} \\
&= \begin{bmatrix} \frac{1}{2} & -\frac{1}{14} \\ -\frac{1}{14} & \frac{1}{84} \end{bmatrix} - \begin{bmatrix} \frac{5}{12} & -\frac{5}{84} \\ -\frac{5}{84} & \frac{5}{504} \end{bmatrix} \\
&= \det \begin{bmatrix} \frac{1}{12} & -\frac{1}{84} \\ -\frac{1}{84} & \frac{1}{504} \end{bmatrix}
\end{aligned}$$

Recall that a matrix Δ is positive definite if and only if all its eigenvalues are strictly positive. For a 2×2 matrix, the eigenvalues are the solutions to:

$$\lambda^2 - tr\left(\Delta\right) \cdot \lambda + \det\left(\Delta\right) = 0$$

Which, in this case is:

$$\lambda^2 - \left(\frac{43}{504}\right)\lambda + \frac{1}{42336} = 0$$

Or,

$$\lambda \in \{0.0848, 0.0003\}$$

Both of which are positive. Therefore, Δ is positive definite, and we can say that, in matrix sense,

$$\widehat{Var\left(\hat{\beta}_{OLS}\right)} > \widehat{Var\left(\hat{\beta}_{FGLS}\right)}$$

3. Suppose we have n i.i.d. observations on (Y_i, X_i) from the $K-$variate linear regression model

$$Y_i = X_i\beta + \varepsilon_i$$

where $E\left(\varepsilon_i | x_i\right) = 0$ and $Var\left(\varepsilon_i | x_i\right) = \sigma^2$. Under what conditions is the following estimator of σ^2

$$\hat{\sigma}^2 = \frac{1}{n-K}\sum_{i=1}^{n}\left(Y_i - X_i\hat{\beta}_{OLS}\right)^2$$

consistent and asymptotically normal? Derive the asymptotic variance and provide a consistent estimator for it.

Solution First, throughout this problem, I will make use of the fact that

$$\sigma^2 = Var\left(\varepsilon_i | X_i\right) = E\left[\varepsilon_i^2 \middle| X_i\right] - \underbrace{E\left[\varepsilon_i | X_i\right]}_{=0} E\left[\varepsilon_i | X_i\right] = E\left[\varepsilon_i^2 \middle| X_i\right]$$

Recall that

$$\hat{\sigma}_n^2 = \frac{1}{n-k}\hat{\varepsilon}'\hat{\varepsilon} = \frac{1}{n-k}\varepsilon' M_X \varepsilon$$

Expanding this out, we have:

$$\begin{aligned}
\hat{\sigma}_n^2 &= \frac{1}{n-k}\varepsilon'\left(I - X\left(X'X\right)^{-1}X'\right)\varepsilon \\
&= \frac{1}{n-k}\left(\varepsilon'\varepsilon - \varepsilon' X\left(X'X\right)^{-1}X'\varepsilon\right) \\
&= \frac{1}{n-k}\left(\varepsilon'\varepsilon - \left(X'\varepsilon\right)'\left(X'X\right)^{-1}\left(X'\varepsilon\right)\right)
\end{aligned}$$

In order to do any proofs using asymptotic theory, it is necessary first to put everything in terms of sample averages:

$$\begin{aligned}
\hat{\sigma}_n^2 &= \frac{1}{n-k}\left(n\cdot\frac{1}{n}\varepsilon'\varepsilon - n\left(\frac{1}{n}X'\varepsilon\right)'\left(\frac{1}{n}X'X\right)^{-1}\left(\frac{1}{n}X'\varepsilon\right)\right) \\
&= \frac{n}{n-k}\left(\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n}X_i'\varepsilon_i\right)'\left(\frac{1}{n}\sum_{i=1}^{n}X_i'X_i\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}X_i'\varepsilon_i\right)\right) \qquad (1)
\end{aligned}$$

This question asks us to derive conditions that allow us to have the following two results:

$$p\lim\hat{\sigma}_n^2 = \sigma^2$$

And

$$\sqrt{n}\left(\hat{\sigma}_n^2 - \sigma^2\right) \xrightarrow{d} N\left(0, AV\left(\hat{\sigma}_n^2\right)\right)$$

Proceeding first with showing $p\lim\hat{\sigma}_n^2 = \sigma^2$, first recall the following results:

$$\begin{aligned}
p\lim a_n b_n &= \left(p\lim a_n\right)\left(p\lim b_n\right) \\
p\lim\left(a_n + b_n\right) &= p\lim a_n + p\lim b_n \\
p\lim a_n^{-1} &= \left(p\lim a_n\right)^{-1},\ \det\left(p\lim a_n\right) \neq 0
\end{aligned}$$

The first two hold because the multiplication and addition operations, respectively, are continuous functions. (The first two are also results of Slutsky's theorem.) The third holds because the inverse operation for matrices is continuous except at the point where the matrix is not invertible.

This gives us (from (1)):

$$\begin{aligned}
p\lim \hat{\sigma}_n^2 &= p\lim\left[\frac{n}{n-k}\left(\frac{1}{n}\sum_{i=1}^n \varepsilon_i^2 - \left(\frac{1}{n}\sum_{i=1}^n X_i'\varepsilon_i\right)'\left(\frac{1}{n}\sum_{i=1}^n X_i'X_i\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^n X_i'\varepsilon_i\right)\right)\right] \\
&= \left(\underbrace{p\lim\frac{n}{n-k}}_{(2)}\right)\left(\begin{array}{c}\underbrace{p\lim\frac{1}{n}\sum_{i=1}^n \varepsilon_i^2}_{(3)} - \left(\underbrace{p\lim\frac{1}{n}\sum_{i=1}^n X_i'\varepsilon_i}_{(4)}\right)' \cdot \\ \left(\underbrace{p\lim\frac{1}{n}\sum_{i=1}^n X_i'X_i}_{(5)}\right)^{-1}\left(\underbrace{p\lim\frac{1}{n}\sum_{i=1}^n X_i'\varepsilon_i}_{(6)}\right)\end{array}\right)
\end{aligned}$$

Since we can think of a sequence of degenerate random variables as just a nonstochastic sequence, we have that, for (2):

$$p\lim\frac{n}{n-k} = \lim\frac{n}{n-k} = 1 \tag{2'}$$

Next, consider (3):

$$p\lim\frac{1}{n}\sum_{i=1}^n \varepsilon_i^2$$

Here, we clearly want to use the weak law of large numbers. We know that, by assumption, $\{Y_i, X_i\}$ i.i.d. implies that ε_i are i.i.d.. this gives us that ε_i^2 are i.i.d.. In order to use the weak law of large numbers, though, we must also assume:

Assumption 1 $E\left[\left|\varepsilon_i^2\right|\right] < +\infty$.

In which case, we get:

$$\begin{aligned}
p\lim\frac{1}{n}\sum_{i=1}^n \varepsilon_i^2 &= E\left[\varepsilon_i^2\right] \\
&= E\left[E\left[\left.\varepsilon_i^2\right| X_i\right]\right] \\
&= E\left[\sigma^2\right] = \sigma^2
\end{aligned} \tag{3'}$$

Where the second-to-last equality holds by assumption.

Proceeding to deal with (4) (and (6)):

$$p\lim\frac{1}{n}\sum_{i=1}^n X_i'\varepsilon_i$$

By the logic exercised in the analysis of (3), we know that since $\{Y_i, X_i\}$ are i.i.d., it follows that $\{\varepsilon_i, X_i\}$ are i.i.d. and hence $X_i'\varepsilon_i$ are i.i.d.. In order to exercise the weak law of large numbers, we must assume:

Assumption 2 $E\left[\left|X_i'\varepsilon_i\right|\right] < +\infty$

This gives us

$$\begin{aligned} p\lim \frac{1}{n}\sum_{i=1}^{n} X_i'\varepsilon_i &= E\left[X_i'\varepsilon_i\right] \\ &= E\left[E\left[X_i'\varepsilon_i \middle| X_i\right]\right] \\ &= E\left[X_i' E\left[\varepsilon_i \middle| X_i\right]\right] \\ &= E\left[X_i' \cdot 0\right] = 0 \end{aligned} \tag{4'}$$

Where the second-to-last equality once again holds by assumption.

This part of the question is actually complete, but for the sake of clarity, I will also characterize the behavior of (5):

$$p\lim \frac{1}{n}\sum_{i=1}^{n} X_i' X_i$$

Exercising the same logic, it is clear that $X_i'X_i$ are i.i.d.. Thus, if we assume

Assumption 3 $E\left[|X_i'X_i|\right] < +\infty$ and positive definite. (Hence invertible)

Then we have:

$$p\lim \frac{1}{n}\sum_{i=1}^{n} X_i' X_i = E\left[X_i'X_i\right] \tag{5'}$$

By the weak law of large numbers.

Putting all this together, we get:

$$\begin{aligned} p\lim \hat{\sigma}_n^2 &= 1\left(\sigma^2 + (0)'\left(E\left[X_i'X_i\right]\right)^{-1}(0)\right) \\ &= \sigma^2 \end{aligned}$$

Indeed, $\hat{\sigma}_n^2$ is consistent for σ^2 given the following assumptions:

$$\begin{aligned} E\left[\left|\varepsilon_i^2\right|\right] &< +\infty \\ E\left[|X_i'\varepsilon_i|\right] &< +\infty \\ E\left[|X_i'X_i|\right] &< +\infty \end{aligned}$$

It can be easily shown that assumption 1 is redundant, however:

$$E\left[\left|\varepsilon_i^2\right|\right] = E\left[\varepsilon_i^2\right] = E\left[E\left[\varepsilon_i^2 \middle| X_i\right]\right] = E\left[\sigma^2\right] = \sigma^2 < +\infty$$

The next part of the question asks us to find the conditions necessary for the asymptotic normality of $\hat{\sigma}_n^2$. In order to do so, we must first solve for $\sqrt{n}\left(\hat{\sigma}_n^2 - \sigma^2\right)$

$$\begin{aligned} &\sqrt{n}\left(\hat{\sigma}_n^2 - \sigma^2\right) \\ =\ &\sqrt{n}\left[\frac{n}{n-k}\left(\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n}X_i'\varepsilon_i\right)'\left(\frac{1}{n}\sum_{i=1}^{n}X_i'X_i\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}X_i'\varepsilon_i\right)\right) - \sigma^2\right] \\ =\ &\sqrt{n}\left[\begin{array}{c} \frac{n}{n-k}\left(\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2 - \sigma^2\right) + \frac{k}{n-k}\sigma^2 \\ -\frac{n}{n-k}\left(\frac{1}{n}\sum_{i=1}^{n}X_i'\varepsilon_i\right)'\left(\frac{1}{n}\sum_{i=1}^{n}X_i'X_i\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}X_i'\varepsilon_i\right) \end{array}\right] \end{aligned}$$

$$= \underbrace{\frac{\sqrt{n}k}{n-k}}_{(7)}\sigma^2 + \underbrace{\frac{n}{n-k}}_{(8)} \left[\begin{array}{c} \underbrace{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2 - \sigma^2\right)}_{(9)} \\ -\underbrace{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}X_i'\varepsilon_i\right)'}_{(10)} \left(\underbrace{\frac{1}{n}\sum_{i=1}^{n}X_i'X_i}_{(11)}\right)^{-1} \left(\underbrace{\frac{1}{n}\sum_{i=1}^{n}X_i'\varepsilon_i}_{(12)}\right) \end{array} \right]$$

Since we can think of a sequence of degenerate random variables as just a sequence of nonstochastic terms, we have for (7) and (8) :

$$\begin{aligned} p\lim \frac{\sqrt{n}k}{n-k}\sigma^2 &= \lim \frac{\sqrt{n}k}{n-k}\sigma^2 = 0 \\ p\lim \frac{n}{n-k} &= \lim \frac{n}{n-k} = 1 \end{aligned}$$

Next, working on (9), recall the assumption,

$$E\left[\varepsilon_i^2\right] = E\left[E\left[\varepsilon_i^2\middle| X_i\right]\right] = E\left[\sigma^2\right] = \sigma^2$$

Noting that

$$Var\left(\varepsilon_i^2\right) = E\left[\varepsilon_i^4\right] - \left(E\left[\varepsilon_i^2\right]\right)^2 = E\left[\varepsilon_i^4\right] - \left(\sigma^2\right)^2$$

If we assume

Assumption 4 $E\left[\varepsilon_i^4\right] < +\infty$

We have $Var\left(\varepsilon_i^2\right) < +\infty$. Then, with the i.i.d. assumption, we can invoke the central limit theorem to give us:

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2 - \sigma^2\right) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2 - E\left[\varepsilon_i^2\right]\right) \xrightarrow{d} N\left(0, E\left[\varepsilon_i^4\right] - \left(\sigma^2\right)^2\right)$$

If we focus now on (10), we need to make use of the facts that

$$E\left[X_i'\varepsilon_i\right] = E\left[E\left[X_i'\varepsilon_i\middle| X_i\right]\right] = E\left[X_i'E\left[\varepsilon_i\middle| X_i\right]\right] = E\left[X_i'\cdot 0\right] = 0$$

And $X_i'\varepsilon_i$ are i.i.d. Further, by assumption 3, we have that $E\left[X_i'X_i\right] \leq E\left[\left|X_i'X_i\right|\right] < +\infty$, which gives us:

$$\begin{aligned} Var\left(X_i'\varepsilon_i\right) &= E\left[X_i'\varepsilon_i\left(X_i'\varepsilon_i\right)'\right] - \underbrace{E\left[X_i'\varepsilon_i\right]}_{=0}E\left[X_i'\varepsilon_i\right]' \\ &= E\left[X_i'\varepsilon_i\varepsilon_i'X_i\right] = E\left[X_i'X_i\varepsilon_i^2\right] \\ &= E\left[E\left[X_i'X_i\varepsilon_i^2\middle| X_i\right]\right] = E\left[X_i'X_iE\left[\varepsilon_i^2\middle| X_i\right]\right] \\ &= \sigma^2 E\left[X_i'X_i\right] < \infty \end{aligned}$$

Where the third equality holds since ε_i is a scalar. (And thus $\varepsilon_i = \varepsilon_i'$). Therefore, we can invoke the central limit theorem to give us:

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}X_i'\varepsilon_i\right) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}X_i'\varepsilon_i - E\left[X_i'\varepsilon_i\right]\right) \xrightarrow{d} N\left(0, \sigma^2 E\left[X_i'X_i\right]\right)$$

From the analysis of the consistency of $\hat{\sigma}_n^2$, we know that for (11) :

$$\left(p\lim \frac{1}{n}\sum_{i=1}^{n} X_i'X_i\right)^{-1} = \left(E\left[X_i'X_i\right]\right)^{-1}$$

Where I used the fact that $X_i'X_i$ are i.i.d., the Mann-Wald theorem, assumption 3, and the weak law of large numbers.

Finally, for (12), we have by the weak law of large numbers:

$$p\lim \frac{1}{n}\sum_{i=1}^{n} X_i'\varepsilon_i = E\left[X_i'\varepsilon_i\right] = 0$$

It will require a few steps to put this all together. First, by Slutsky's theorem, we have that

$$\underbrace{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i'\varepsilon_i\right)'}_{\xrightarrow{d} N\left(0,\sigma^2 E\left[X_i'X_i\right]\right)}\underbrace{\left(\frac{1}{n}\sum_{i=1}^{n} X_i'X_i\right)^{-1}}_{\xrightarrow{p}\left(E\left[X_i'X_i\right]\right)^{-1}}\underbrace{\left(\frac{1}{n}\sum_{i=1}^{n} X_i'\varepsilon_i\right)}_{\xrightarrow{p}0} \xrightarrow{d} 0$$

Which implies convergence in probability to zero as well. (Since convergence in distribution and convergence in probability are equivalent when the limit is a constant.)

Next, we have by Slutsky's theorem:

$$\begin{aligned}&\underbrace{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i^2 - \sigma^2\right)}_{\xrightarrow{d} N\left(0,E\left[\varepsilon_i^4\right]-(\sigma^2)^2\right)} - \underbrace{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i'\varepsilon_i\right)'\left(\frac{1}{n}\sum_{i=1}^{n} X_i'X_i\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} X_i'\varepsilon_i\right)}_{\xrightarrow{p}0}\\ &\xrightarrow{d} N\left(0, E\left[\varepsilon_i^4\right] - \left(\sigma^2\right)^2\right)\end{aligned}$$

Finally, by using Slutsky's theorem twice, we have:

$$\begin{aligned}&\underbrace{\frac{\sqrt{n}k}{n-k}\sigma^2}_{\xrightarrow{p}0} + \underbrace{\frac{n}{n-k}}_{\xrightarrow{p}1}\underbrace{\left[\begin{array}{c}\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2-\sigma^2\right)\\ -\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i'\varepsilon_i\right)'\left(\frac{1}{n}\sum_{i=1}^{n} X_i'X_i\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} X_i'\varepsilon_i\right)\end{array}\right]}_{\xrightarrow{d} N\left(0,E\left[\varepsilon_i^4\right]-(\sigma^2)^2\right)}\\ &\xrightarrow{d} N\left(0, E\left[\varepsilon_i^4\right] - \left(\sigma^2\right)^2\right)\end{aligned}$$

And therefore,

$$\sqrt{n}\left(\hat{\sigma}_n^2 - \sigma^2\right) \xrightarrow{d} N\left(0, E\left[\varepsilon_i^4\right] - \left(\sigma^2\right)^2\right)$$

Where the only additional assumption was that $E\left[\varepsilon_i^4\right] < +\infty$

Here, we have that $AV\left(\hat{\sigma}_n^2\right) = E\left[\varepsilon_i^4\right] - \left(\sigma^2\right)^2$. The final part of this question asks us to derive $\widehat{AV\left(\hat{\sigma}_n^2\right)}$, a consistent estimator for $AV\left(\hat{\sigma}_n^2\right)$. I claim that

$$\widehat{AV\left(\hat{\sigma}_n^2\right)} = \frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^4 - \left(\hat{\sigma}_n^2\right)^2$$

is a consistent estimator, which I will now establish. I begin by noting that

$$\begin{aligned}\hat{\varepsilon}_i &= Y_i - X_i\hat{\beta}_n \\ &= X_i\beta + \varepsilon_i - X_i\hat{\beta}_n \\ &= \varepsilon_i + X_i\left(\beta - \hat{\beta}_n\right)\end{aligned}$$

And therefore,

$$\hat{\varepsilon}_i^4 = \varepsilon_i^4 + 4\varepsilon_i^3 X_i\left(\beta - \hat{\beta}_n\right) + 6\varepsilon_i^2\left[X_i\left(\beta - \hat{\beta}_n\right)\right]^2 + 4\varepsilon_i\left[X_i\left(\beta - \hat{\beta}_n\right)\right]^3 + \left[X_i\left(\beta - \hat{\beta}_n\right)\right]^4$$

Or,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^4 &= \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^4 + \frac{4}{n}\sum_{i=1}^{n}\varepsilon_i^3 X_i\left(\beta - \hat{\beta}_n\right) + \frac{6}{n}\sum_{i=1}^{n}\varepsilon_i^2\left[X_i\left(\beta - \hat{\beta}_n\right)\right]^2 \\ &\quad + \frac{4}{n}\sum_{i=1}^{n}\varepsilon_i\left[X_i\left(\beta - \hat{\beta}_n\right)\right]^3 + \frac{1}{n}\sum_{i=1}^{n}\left[X_i\left(\beta - \hat{\beta}_n\right)\right]^4\end{aligned}$$

This complicated expression can be broken down with the help of the following assumptions:

Assumption 5 $\varepsilon_i^3 = O_p(1)\,,\ \varepsilon_i^2 = O_p(1)\,,\ \varepsilon_i = O_p(1)\,,\ X_i = O_p(1)$.

Assumption 6 $E\left[\left|\varepsilon_i^4\right|\right] = E\left[\varepsilon_i^4\right] < +\infty$

By the weak law of large numbers, assumption 6, and the i.i.d. data assumption, we have that $\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^4 = E\left[\varepsilon_i^4\right] + o_p(1)$. Also, since $\hat{\beta}_n$ is consistent for β, we have that $\hat{\beta}_n = \beta + o_p(1)$. Recall the following results:

$$\begin{aligned}O_p(1)\cdot o_p(1) &= o_p(1) \\ o_p(1)\cdot o_p(1) &= o_p(1) \\ o_p(1) + o_p(1) &= o_p(1) \\ -o_p(1) &= o_p(1)\end{aligned}$$

This gives us:

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^4 &= E\left[\varepsilon_i^4\right] + o_p(1) + \frac{4}{n}\sum_{i=1}^{n}O_p(1)\left[O_p(1)\left(\beta - \left(\beta + o_p(1)\right)\right)\right] \\ &\quad + \frac{6}{n}\sum_{i=1}^{n}O_p(1)\left[O_p(1)\left(\beta - \left(\beta + o_p(1)\right)\right)\right]^2 \\ &\quad + \frac{4}{n}\sum_{i=1}^{n}O_p(1)\left[O_p(1)\left(\beta - \left(\beta + o_p(1)\right)\right)\right]^3 \\ &\quad + \frac{1}{n}\sum_{i=1}^{n}\left[O_p(1)\left(\beta - \left(\beta + o_p(1)\right)\right)\right]^4\end{aligned}$$

$$\begin{aligned}&= E\left[\varepsilon_i^4\right] + o_p(1) + \frac{4}{n}\sum_{i=1}^{n}O_p(1)\left[O_p(1)\,o_p(1)\right] + \frac{6}{n}\sum_{i=1}^{n}O_p(1)\left[O_p(1)\,o_p(1)\right]^2 \\ &\quad + \frac{4}{n}\sum_{i=1}^{n}O_p(1)\left[O_p(1)\,o_p(1)\right]^3 + \frac{1}{n}\sum_{i=1}^{n}\left[O_p(1)\,o_p(1)\right]^4\end{aligned}$$

$$
\begin{aligned}
&= E\left[\varepsilon_i^4\right] + o_p(1) + \frac{4}{n}\sum_{i=1}^{n} O_p(1)\, o_p(1) + \frac{6}{n}\sum_{i=1}^{n} O_p(1)\, o_p(1)\, o_p(1) \\
&\quad + \frac{4}{n}\sum_{i=1}^{n} O_p(1)\, o_p(1)\, o_p(1)\, o_p(1) + \frac{1}{n}\sum_{i=1}^{n} o_p(1)\, o_p(1)\, o_p(1)\, o_p(1) \\
&= E\left[\varepsilon_i^4\right] + o_p(1) + \frac{4}{n}\sum_{i=1}^{n} o_p(1) + \frac{6}{n}\sum_{i=1}^{n} o_p(1) + \frac{4}{n}\sum_{i=1}^{n} o_p(1) + \frac{1}{n}\sum_{i=1}^{n} o_p(1)
\end{aligned}
$$

$$
\begin{aligned}
&= E\left[\varepsilon_i^4\right] + o_p(1) + o_p(1) + o_p(1) + o_p(1) + o_p(1) \\
&= E\left[\varepsilon_i^4\right] + o_p(1)
\end{aligned}
$$

That is,

$$
p\lim \frac{1}{n}\sum_{i=1}^{n} \hat{\varepsilon}_i^4 = E\left[\varepsilon_i^4\right]
$$

Finally, as proved earlier in this question,

$$
p\lim \hat{\sigma}_n^2 = \sigma^2
$$

So, by the Mann-Wald theorem,

$$
p\lim \left(\hat{\sigma}_n^2\right)^2 = \left(\sigma^2\right)^2
$$

Invoking Slutsky's theorem:

$$
\begin{aligned}
p\lim \widehat{AV\left(\hat{\sigma}_n^2\right)} &= p\lim \frac{1}{n}\sum_{i=1}^{n} \hat{\varepsilon}_i^4 - p\lim \left(\hat{\sigma}_n^2\right)^2 \\
&= E\left[\varepsilon_i^4\right] - \left(\sigma^2\right)^2 = AV\left(\hat{\sigma}_n^2\right)
\end{aligned}
$$

In particular, $\widehat{AV\left(\hat{\sigma}_n^2\right)}$ is consistent for $AV\left(\hat{\sigma}_n^2\right)$.

4. Suppose we have n i.n.i.d. observations on (Y_i, X_i) from the $K-$variate linear regression model

$$Y_i = X_i\beta + \varepsilon_i$$

where X_i is fixed across repeated samples and $E(\varepsilon_i) = 0$ and $Var(\varepsilon_i) = \sigma^2$ and ε_i is independent of ε_j for $i \neq j$. Derive the asymptotic properties of $\hat{\beta}_{OLS}$. Make sure to state all necessary assumptions and theorems.

Solution I will take the comment "X_i is fixed across repeated samples" to mean that the X_i are fixed regressors. One implication of this is that the Y_i are not *i.i.d.* as can be quickly seen by noting that $E[Y_i] = E[X_i\beta + \varepsilon_i] = X_i\beta$, which, in general is not equal to $E[Y_j]$ for $i \neq j$. As a result, we will need to use more powerful versions of the weak law of large numbers and the central limit theorem. Recall the following theorems from the appendices of lecture notes 7:

Theorem 6.7. (Markov) Let $\{X_i\}$ be a sequence of independent random variables with finite means $\mu_i \equiv E[X_i]$. If for some $\delta > 0$, $\sum_{i=1}^{\infty} \frac{E\left[|X_i - \mu_i|^{1+\delta}\right]}{i^{1+\delta}} < +\infty$, then

$$\bar{X}_n - \bar{\mu}_n \overset{a.s.}{\to} 0$$

We may use this theorem if we make the following additional assumptions:

Assumption 1 $E\left[|X_{ik}\varepsilon_i|^{1+\delta}\right] < \Delta < +\infty$ for all $i = 1, \ldots, n$ and for all $k = 1, \ldots, K$, where Δ is simply some uniform upper bound.

Assumption 2 $E\left[|X_{ik}^2|^{1+\delta}\right] < \Delta < \infty$ for all $i = 1, \ldots, n$ and for all $k = 1, \ldots, K$, and

Assumption 3 $E\left[\frac{1}{n}X'X\right]$ is uniformly positive definite.

Note that the assumption of independence is already satisfied since the ε_i are independent and they are the only stochastic elements of this model. Further, we also have $E[X_i'\varepsilon_i] = X_i'E[\varepsilon_i] = X_i' \cdot 0 = 0$.

The other theorem we will want to invoke is:

Theorem 6.23. (Liapounov) Suppose that $\{X_i\}_{i=1}^{\infty}$ is a sequence of independent RV's such that $E[X_i] = \mu_i < +\infty$, $Var(X_i) = \sigma_i^2 < +\infty$ and $E\left[|X_i - \mu_i|^{2+\delta}\right] < \Delta < +\infty$ for some $\delta > 0$ and for all i. If $\bar{\sigma}_n^2 > \delta' > 0$ for all n sufficiently large, then

$$Z_n \equiv \sqrt{n}\frac{\bar{X}_n - \bar{\mu}_n}{\sqrt{\bar{\sigma}_n^2}} \overset{d}{\to} N(0,1) \tag{1}$$

Remark I will rewrite (1) in terms of something with which we are more familiar:

$$\sqrt{n}\left(\bar{X}_n - \bar{\mu}_n\right) \overset{d}{\to} N\left(0, \lim \bar{\sigma}_n^2\right)$$

In order to use this theorem, we will need to make the additional assumptions:

Assumption 4 $E\left[|X_i - \mu_i|^{2+\delta}\right] < \Delta < +\infty$ for some $\delta > 0$ and for all i.

Assumption 5 $\bar{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}\sigma_i^2 > \delta' > 0$ for all n sufficiently large for some δ'. (That is, X_i has positive variance almost always.)

With this theoretical baggage out of the way, let us proceed to derive the asymptotic properties of $\hat{\beta}_n^{OLS}$.

As usual, write $\hat{\beta}_n^{OLS}$ in terms of sample averages:

$$\begin{aligned} \hat{\beta}_n^{OLS} &= \beta + (X'X)^{-1} X'\varepsilon \\ &= \beta + \left(\frac{1}{n}X'X\right)^{-1}\left(\frac{1}{n}X'\varepsilon\right) \\ &= \beta + \left(\frac{1}{n}\sum_{i=1}^{n} X_i'X_i\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} X_i'\varepsilon_i\right) \end{aligned} \tag{2}$$

First, let us examine whether or not $p\lim \hat{\beta}_n^{OLS} = \beta$. (That is, whether or not $\hat{\beta}_n^{OLS}$ is a consistent estimator of β.) By Slutsky's theorem and the Mann-Wald theorem, we have the result that

$$p\lim \hat{\beta}_n^{OLS} = \beta + \left(\underbrace{p\lim \frac{1}{n}\sum_{i=1}^{n} X_i'X_i}_{(1)}\right)^{-1}\left(\underbrace{p\lim \frac{1}{n}\sum_{i=1}^{n} X_i'\varepsilon_i}_{(2)}\right)$$

For (2), recognize that $X_i'X_i$ is a nonstochastic sequence. Therefore,

$$p\lim \frac{1}{n}\sum_{i=1}^{n} X_i'X_i = \lim \frac{1}{n}\sum_{i=1}^{n} X_i'X_i$$

In order to proceed, we need to assume that this limit exists (and is invertible):

Assumption 6 $\lim \frac{1}{n}\sum_{i=1}^{n} X_i'X_i \equiv G < +\infty$ and G is invertible.

For (2), we must use Markov's law of large numbers theorem from above. First, define $\mu_i = E\left[X_i'\varepsilon_i\right] = X_i'E\left[\varepsilon_i\right] = X_i' \cdot 0 = 0 < +\infty$. Then we have that:

$$0 = p\lim\left(\frac{1}{n}\sum_{i=1}^{n} X_i'\varepsilon_i - \frac{1}{n}\sum_{i=1}^{n}\mu_i\right) = p\lim\left(\frac{1}{n}\sum_{i=1}^{n} X_i'\varepsilon_i\right)$$

Putting this together, we have:

$$p\lim \hat{\beta}_n^{OLS} = \beta + \left(\lim \frac{1}{n}\sum_{i=1}^{n} X_i'X_i\right)^{-1} \cdot 0 = \beta$$

That is, $\hat{\beta}_n^{OLS}$ is consistent for β. Next, proceeding to demonstrate normality. Rewriting (1) gives us:

$$\begin{aligned} \hat{\beta}_n^{OLS} - \beta &= \left(\frac{1}{n}\sum_{i=1}^{n} X_i'X_i\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} X_i'\varepsilon_i\right) \\ \sqrt{n}\left(\hat{\beta}_n^{OLS} - \beta\right) &= \left(\frac{1}{n}\sum_{i=1}^{n} X_i'X_i\right)^{-1}\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i'\varepsilon_i\right) \end{aligned}$$

From the analysis of consistency, we saw that

$$p\lim \frac{1}{n}\sum_{i=1}^{n} X_i'X_i = \lim \frac{1}{n}\sum_{i=1}^{n} X_i'X_i \equiv G < +\infty$$

And by the Mann-Wald theorem,

$$\begin{aligned} p\lim\left(\frac{1}{n}\sum_{i=1}^{n}X_i'X_i\right)^{-1} &= \left(p\lim\frac{1}{n}\sum_{i=1}^{n}X_i'X_i\right)^{-1} \\ &= G^{-1} \end{aligned}$$

Next, note that

$$\mu_i = E\left[X_i'\varepsilon_i\right] = X_i'E\left[\varepsilon_i\right] = X_i' \cdot 0 = 0$$

This allows us to use Liapounov's central limit theorem for i.n.i.d. data:

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}X_i'\varepsilon_i - \frac{1}{n}\sum_{i=1}^{n}\mu_i\right) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}X_i'\varepsilon_i\right) \xrightarrow{d} N\left(0, \lim\frac{1}{n}Var\left(X_i'\varepsilon_i\right)\right)$$

Where, since $Var\left(X_i'\varepsilon_i\right) = X_i'Var\left(\varepsilon_i\right)X_i = \sigma^2 X_i'X_i$, we have that

$$\lim\frac{1}{n}Var\left(X_i'\varepsilon_i\right) = \sigma^2\lim\frac{1}{n}\sum_{i=1}^{n}X_i'X_i = \sigma^2 G < +\infty$$

Thus,

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}X_i'\varepsilon_i\right) \xrightarrow{d} N\left(0, \sigma^2 G\right)$$

Using Slutsky's theorem and putting this all together, we have:

$$\sqrt{n}\left(\hat{\beta}_n^{OLS} - \beta\right) \xrightarrow{d} N\left(0, \sigma^2 G^{-1}GG^{-1}\right)$$

Or

$$\sqrt{n}\left(\hat{\beta}_n^{OLS} - \beta\right) \xrightarrow{d} N\left(0, \sigma^2 G^{-1}\right)$$

In particular, $\hat{\beta}_n^{OLS}$ is asymptotically normal.

5. A groupwise heteroskedastic regression has

$$Y_i = X_i\beta + \varepsilon_i$$

where

$$E\left(\varepsilon_i | X_i\right) = 0$$

The n independent observations are grouped into G groups each n_g observations. The slope vector β is the same in all groups but within group g,

$$Var\left(\varepsilon_i | x_i\right) = \sigma_g^2$$

a. Derive the GLS, FGLS, and ML estimator of β.

Solution Here, we have

$$Var\left(\varepsilon | X\right) = \begin{bmatrix} \sigma_1^2 & & & & & & \\ & \ddots & & & & & \\ & & \sigma_1^2 & & & & \\ & & & \ddots & & & \\ & & & & \sigma_G^2 & & \\ & & & & & \ddots & \\ & & & & & & \sigma_G^2 \end{bmatrix} \equiv \Omega$$

Where I will assume that all the off-diagonal elements are zero (the case of pure heteroskedasticity).

First, computing $\hat{\beta}_{GLS}$:

$$\begin{aligned} \hat{\beta}_{GLS} &= \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}Y \\ &= \left(\begin{bmatrix} X_1' & \cdots & X_{n_1}' & \cdots & X_{n_1+\cdots+n_{G-1}+1}' & \cdots & X_n' \end{bmatrix} \cdot \Omega^{-1} \cdot \begin{bmatrix} X_1 \\ \vdots \\ X_{n_1} \\ \vdots \\ X_{n_1+\cdots+n_{G-1}+1} \\ \vdots \\ X_{n_G} \end{bmatrix} \right) \cdot \\ & \quad \begin{bmatrix} X_1' & \cdots & X_{n_1}' & \cdots & X_{n_1+\cdots+n_{G-1}+1}' & \cdots & X_n' \end{bmatrix} \cdot \Omega^{-1} \cdot \begin{bmatrix} Y_1 \\ \vdots \\ Y_{n_1} \\ \vdots \\ Y_{n_1+\cdots+n_{G-1}+1} \\ \vdots \\ Y_{n_G} \end{bmatrix} \end{aligned}$$

$$
\begin{aligned}
&= \left(\sum_{i=1}^{n_1}\frac{1}{\sigma_1^2}X_i'X_i+\cdots+\sum_{i=n_1+\cdots+n_{G-1}+1}^{n}\frac{1}{\sigma_G^2}X_i'X_i\right)^{-1}\cdot \\
&\quad \left(\sum_{i=1}^{n_1}\frac{1}{\sigma_1^2}X_i'Y_i+\cdots+\sum_{i=n_1+\cdots+n_{G-1}+1}^{n}\frac{1}{\sigma_G^2}X_i'Y_i\right) \\
&= \left(\sum_{g=1}^{G}\sum_{i=n_1+\cdots+n_{g-1}+1}^{n_1+\cdots+n_g}\frac{X_i'X_i}{\sigma_g^2}\right)^{-1}\left(\sum_{g=1}^{G}\sum_{i=n_1+\cdots+n_{g-1}+1}^{n_1+\cdots+n_g}\frac{X_i'Y_i}{\sigma_g^2}\right) \qquad (1)
\end{aligned}
$$

If we define

$$
X_g=\begin{bmatrix} X_{n_1+\cdots+n_{g-1}+1} \\ \vdots \\ X_{n_1+\cdots+n_g}\end{bmatrix},\ Y_g=\begin{bmatrix} Y_{n_1+\cdots+n_{g-1}+1} \\ \vdots \\ Y_{n_1+\cdots+n_g}\end{bmatrix}
$$

Then immediately, we get

$$
\begin{aligned}
X_g'X_g &= \sum_{i=n_1+\cdots+n_{g-1}+1}^{n_1+\cdots+n_g} X_i'X_i \\
X_g'Y_g &= \sum_{i=n_1+\cdots+n_{g-1}+1}^{n_1+\cdots+n_g} X_i'Y_i
\end{aligned}
$$

Which we can substitute into (1):

$$
\hat{\beta}_{GLS}=\left(\sum_{g=1}^{G}\frac{X_g'X_g}{\sigma_g^2}\right)^{-1}\left(\sum_{g=1}^{G}\frac{X_g'Y_g}{\sigma_g^2}\right)
$$

The following result for estimating $\hat{\beta}_{FGLS}$ uses the two-step FGLS procedure introduced in question 2, part d:

First, estimate $\hat{\beta}_{OLS}=(X'X)^{-1}X'Y$ (using the entire dataset), from which we can obtain $\hat{\varepsilon}=\begin{bmatrix}\hat{\varepsilon}_1 \\ \vdots \\ \hat{\varepsilon}_G\end{bmatrix}=$ $Y-X\hat{\beta}_{OLS}$, where $\hat{\varepsilon}_g=\begin{bmatrix}\hat{\varepsilon}_{n_1+\cdots+n_{g-1}+1} \\ \vdots \\ \hat{\varepsilon}_{n_1+\cdots+n_g}\end{bmatrix}$.

Using this, estimate $\hat{\sigma}_g^2=\frac{\hat{\varepsilon}_g'\hat{\varepsilon}_g}{n_g-k}$ for $g=1,\ldots,G$ to construct

$$
\hat{\Omega}=\begin{bmatrix}
\hat{\sigma}_1^2 & & & & & \\
& \ddots & & & & \\
& & \hat{\sigma}_1^2 & & & \\
& & & \ddots & & \\
& & & & \hat{\sigma}_G^2 & \\
& & & & & \ddots & \\
& & & & & & \hat{\sigma}_G^2
\end{bmatrix}
$$

From which we can obtain the feasible least squares estimator:

$$\hat{\beta}_{FGLS} = \left(X'\hat{\Omega}^{-1}X\right)^{-1} X'\hat{\Omega}^{-1}Y = \left(\sum_{g=1}^{G} \frac{X_g'X_g}{\hat{\sigma}_g^2}\right)^{-1} \left(\sum_{g=1}^{G} \frac{X_g'Y_g}{\hat{\sigma}_g^2}\right)$$

Proceeding to the last part of the question, since maximum likelihood estimators are only defined for particular distributions, we will need to make an assumption about the distribution of ε_i. I will make the standard assumption that $\varepsilon_i \sim N\left(0, \sigma_g^2\right)$ for $n_1 + \cdots + n_{g-1} < i \leq n_1 + \cdots + n_g$. As usual, we will construct the log-likelihood function beginning with the likelihood function:

$$L\left(\beta, \{\sigma_g\}_{g=1}^{G}; x, y\right) = \prod_{g=1}^{G} \prod_{i=n_1+\cdots+n_{g-1}+1}^{n_1+\cdots+n_g} \frac{1}{\sqrt{2\pi}\sqrt{\sigma_g^2}} \exp\left\{-\frac{(y_i - x_i\beta)^2}{2\sigma_g^2}\right\}$$

Taking logs,

$$\begin{aligned}
\log L &= \sum_{g=1}^{G} \sum_{i=n_1+\cdots+n_{g-1}+1}^{n_1+\cdots+n_g} \left[-\frac{1}{2}\log 2\pi - \frac{1}{2}\log\sigma_g^2 - \frac{(y_i - x_i\beta)^2}{2\sigma_g^2}\right] \\
&= -\frac{n}{2}\log 2\pi - \frac{1}{2}\sum_{g=1}^{G} n_g \log\sigma_g^2 - \frac{1}{2}\sum_{g=1}^{G} \sum_{i=n_1+\cdots+n_{g-1}+1}^{n_1+\cdots+n_g} \frac{(y_i - x_i\beta)^2}{\sigma_g^2}
\end{aligned}$$

Next, taking first order conditions with respect to β to obtain $\hat{\beta}_{MLE}$, we have:

$$(\beta) : -\frac{1}{2}\sum_{g=1}^{G} \sum_{i=n_1+\cdots+n_{g-1}+1}^{n_1+\cdots+n_g} \frac{2\left(y_i - x_i\hat{\beta}_{MLE}\right)(-x_i')}{\hat{\sigma}_g^2} = 0$$

Or,

$$\sum_{g=1}^{G} \sum_{i=n_1+\cdots+n_{g-1}+1}^{n_1+\cdots+n_g} \frac{x_i'y_i - x_i'x_i\hat{\beta}_{MLE}}{\hat{\sigma}_g^2} = 0$$

Where I used the fact that $\left(y_i - x_i\hat{\beta}_{MLE}\right)$ is a scalar in order to establish that $\left(y_i - x_i\hat{\beta}_{MLE}\right)x_i' = x_i'\left(y_i - x_i\hat{\beta}_{MLE}\right)$. Solving for $\hat{\beta}_{MLE}$:

$$\sum_{g=1}^{G} \sum_{i=n_1+\cdots+n_{g-1}+1}^{n_1+\cdots+n_g} \frac{x_i'y_i}{\hat{\sigma}_g^2} = \sum_{g=1}^{G} \sum_{i=n_1+\cdots+n_{g-1}+1}^{n_1+\cdots+n_g} \frac{x_i'x_i}{\hat{\sigma}_g^2}\hat{\beta}_{MLE}$$

$$\hat{\beta}_{MLE} = \left(\sum_{g=1}^{G} \sum_{i=n_1+\ldots n_{g-1}+1}^{n_1+\cdots+n_g} \frac{x_i'x_i}{\hat{\sigma}_g^2}\right)^{-1} \left(\sum_{g=1}^{G} \sum_{i=n_1+\ldots n_{g-1}+1}^{n_1+\cdots+n_g} \frac{x_i'y_i}{\hat{\sigma}_g^2}\right)$$

Since this holds for any particular sample, we have:

$$\hat{\beta}_{MLE} = \left(\sum_{g=1}^{G} \frac{X_g'X_g}{\hat{\sigma}_g^2}\right)^{-1} \left(\sum_{g=1}^{G} \frac{X_g'X_g}{\hat{\sigma}_g^2}\right) = \hat{\beta}_{FGLS}$$

Also, we can derive $\hat{\sigma}_g^2$, which will be useful for part (b) by taking the first order conditions of the log-likelihood function with respect to σ_g^2:

$$(\sigma_g^2) : -\frac{1}{2}\frac{n_g}{\hat{\sigma}_g^2} + \frac{1}{2}\sum_{i=n_1+\ldots n_{g-1}+1}^{n_1+\cdots+n_g} \frac{\left(y_i - x_i\hat{\beta}_{MLE}\right)^2}{\left(\hat{\sigma}_g^2\right)^2} = 0$$

Or (recognizing that $y_i - x_i\hat{\beta}_{FGLS} = \hat{\varepsilon}_i$)

$$\begin{aligned} \sum_{i=n_1+\ldots n_{g-1}+1}^{n_1+\cdots+n_g} \frac{\hat{\varepsilon}_i'\hat{\varepsilon}_i}{\left(\hat{\sigma}_g^2\right)^2} &= \frac{n_g}{\hat{\sigma}_g^2} \\ \frac{\hat{\varepsilon}_g'\hat{\varepsilon}_g}{\left(\hat{\sigma}_g^2\right)^2} &= \frac{n_g}{\hat{\sigma}_g^2} \end{aligned}$$

Finally,

$$\hat{\sigma}_g^2 = \frac{\hat{\varepsilon}_g'\hat{\varepsilon}_g}{n_g}$$

b. Derive the Likelihood Ratio test for group-wise heteroskedasticity and state its asymptotic distribution.

Solution Recall that the Likelihood Ratio test statistic is:

$$LR_0 = 2\left(\log L_U\left(\hat{\theta}_U\right) - \log L_R\left(\hat{\theta}_R\right)\right)$$

Where $L_U,\ L_R,\ \hat{\theta}_U,\ \hat{\theta}_R$ are the likelihood functions and MLEs for the unrestricted and restricted models respectively. In this particular case, we are testing the hypothesis $H_0 : \sigma_1^2 = \cdots = \sigma_G^2 \equiv \sigma^2$. The unrestricted log-likelihood evaluated at the unrestricted MLE function is (by the analysis in part (a)):

$$\begin{aligned} \log L_U\left(\hat{\theta}_U\right) &= -\frac{n}{2}\log 2\pi - \frac{1}{2}\sum_{g=1}^{G} n_g \log \hat{\sigma}_g^2 - \frac{1}{2}\sum_{g=1}^{G}\sum_{i=n_1+\cdots+n_{g-1}+1}^{n_1+\cdots+n_g} \frac{\left(y_i - x_i\hat{\beta}_{MLE}\right)^2}{\hat{\sigma}_g^2} \\ &= -\frac{n}{2}\log 2\pi - \frac{1}{2}\sum_{g=1}^{G} n_g \log \frac{\hat{\varepsilon}_g'\hat{\varepsilon}_g}{n_g} - \frac{1}{2}\sum_{g=1}^{G} \frac{\hat{\varepsilon}_g'\hat{\varepsilon}_g}{\frac{\hat{\varepsilon}_g'\hat{\varepsilon}_g}{n_g}} \\ &= -\frac{n}{2}\log 2\pi - \frac{1}{2}\sum_{g=1}^{G} n_g \log \frac{\hat{\varepsilon}_g'\hat{\varepsilon}_g}{n_g} - \frac{1}{2}\sum_{g=1}^{G} n_g \\ &= -\frac{n}{2}\log 2\pi - \frac{1}{2}\sum_{g=1}^{G} n_g \log \frac{\hat{\varepsilon}_g'\hat{\varepsilon}_g}{n_g} - \frac{n}{2} \end{aligned}$$

The restricted model is the case of conditional homoskedasticity which has the standard MLE estimators:

$$\begin{aligned} \hat{\beta}_{MLE} &= \hat{\beta}_{OLS} \\ \hat{\sigma}_{MLE}^2 &= \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n} \end{aligned}$$

The restricted log-likelihood function evaluated at these parameters is therefore:

$$\begin{aligned} \log L_R\left(\hat{\theta}_R\right) &= -\frac{n}{2}\log 2\pi - \frac{n}{2}\log \hat{\sigma}_{MLE}^2 - \frac{1}{2}\sum_{i=1}^{n} \frac{\left(y_i - x_i\hat{\beta}_{MLE}\right)^2}{\hat{\sigma}^2} \\ &= -\frac{n}{2}\log 2\pi - \frac{n}{2}\log \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n} - \frac{1}{2}\sum_{i=1}^{n} \frac{\hat{\varepsilon}'\hat{\varepsilon}}{\frac{\hat{\varepsilon}'\hat{\varepsilon}}{n}} \\ &= -\frac{n}{2}\log 2\pi - \frac{n}{2}\log \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n} - \frac{n}{2} \end{aligned}$$

Our test statistic is therefore:

$$\begin{aligned}
LR_0 &= 2\left(\log L_U\left(\hat{\theta}_U\right) - \log L_R\left(\hat{\theta}_R\right)\right) \\
&= 2\left(-\frac{n}{2}\log 2\pi - \frac{1}{2}\sum_{g=1}^{G} n_g \log\frac{\hat{\varepsilon}_g'\hat{\varepsilon}_g}{n_g} - \frac{n}{2} - \left(-\frac{n}{2}\log 2\pi - \frac{n}{2}\log\frac{\hat{\varepsilon}'\hat{\varepsilon}}{n} - \frac{n}{2}\right)\right) \\
&= 2\left(\frac{n}{2}\log\frac{\hat{\varepsilon}'\hat{\varepsilon}}{n} - \frac{1}{2}\sum_{g=1}^{G} n_g \log\frac{\hat{\varepsilon}_g'\hat{\varepsilon}_g}{n_g}\right) \\
&= n\log\frac{\hat{\varepsilon}'\hat{\varepsilon}}{n} - \sum_{g=1}^{G} n_g \log\frac{\hat{\varepsilon}_g'\hat{\varepsilon}_g}{n_g}
\end{aligned}$$

It can be shown (see the appendices of lecture notes 8) that

$$LR_0 \xrightarrow{d} \chi^2\left(G-1\right)$$

Where the degrees of freedom $G-1$ are determined by the fact that our null hypothesis makes $G-1$ restrictions.