International School of Economics at TSU
Econometrics 2
Lasha Chochua

Problem Set 6

Instructions: You are encouraged to solve the problems before the recitation. Additionally, you are encouraged to work in groups. It is **not mandatory** to submit solutions unless stated otherwise. However, if you would like to share your solution, I would be happy to review it.

Problem 1: Consider the linear regression model $\mathbf{Y} = \mathbf{X}\beta + \mathbf{e}$, where $\mathbb{E}[\mathbf{e} \mid \mathbf{X}] = \mathbf{0}$ and $\text{Var}[\mathbf{e} \mid \mathbf{X}] = \sigma^2\Sigma$ for some known positive definite matrix Σ.

a. Show that the OLS estimator $\widehat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ remains **unbiased** under heteroskedasticity.

b. Derive the variance of the OLS estimator:

$$\text{Var}[\widehat{\beta} \mid \mathbf{X}] = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\Sigma\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1}.$$

c. Explain why the "standard" covariance estimator $\widehat{V}^0_{\widehat{\beta}} = s^2(\mathbf{X}'\mathbf{X})^{-1}$ is **biased** when $\Sigma \neq I_n$. Under what condition on Σ does this bias vanish?

d. Does heteroskedasticity affect the **unbiasedness** of OLS? Does it affect the **efficiency** of OLS? Explain both answers carefully.

Problem 2: Consider the same setup as Problem 1. The GLS estimator is defined as $\widetilde{\beta}_{gls} = (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}\mathbf{Y}$.

a. Define the transformed model $\tilde{\mathbf{Y}} = \Sigma^{-1/2}\mathbf{Y}$, $\tilde{\mathbf{X}} = \Sigma^{-1/2}\mathbf{X}$, $\tilde{\mathbf{e}} = \Sigma^{-1/2}\mathbf{e}$. Show that $\text{Var}[\tilde{\mathbf{e}} \mid \mathbf{X}] = \sigma^2 I_n$.

b. Using part (a), show that applying OLS to the transformed model yields the GLS estimator:

$$\widetilde{\beta}_{gls} = (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{Y}} = (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}\mathbf{Y}.$$

c. Derive $\text{Var}[\widetilde{\beta}_{gls} \mid \mathbf{X}] = \sigma^2(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}$.

d. Show that when $\Sigma = I_n$ (homoskedasticity), GLS reduces to OLS.

Problem 3: Consider the simple regression $Y_i = \beta_0 + \beta_1 X_i + e_i$ with n observations. Suppose that the true error variance is $\text{Var}[e_i \mid X_i] = \sigma_i^2$ (heteroskedastic).

a. Write down the "sandwich" form of $\text{Var}[\hat{\beta} \mid \mathbf{X}]$ for this model. Express the middle matrix $\mathbf{X}'D\mathbf{X}$ in terms of sums involving X_i and σ_i^2.

b. Suppose $\sigma_i^2 = \sigma^2 X_i^2$ (variance proportional to X_i^2). Explain in words why the classical standard error $s(\hat{\beta}_1) = s\sqrt{[(\mathbf{X}'\mathbf{X})^{-1}]_{22}}$ will generally be **incorrect** in this case.

c. The HC0 estimator replaces e_i^2 with $\hat{e}_i^2$ in the sandwich formula. Why is $\hat{e}_i^2$ a biased estimator of σ_i^2? In which direction is the bias – does $\mathbb{E}[\hat{e}_i^2 \mid \mathbf{X}]$ overestimate or underestimate σ_i^2?

d. The HC1 estimator corrects HC0 by the factor $\frac{n}{n-k}$. Explain the intuition behind this degrees-of-freedom correction. Why is it analogous to dividing by $n-k$ instead of n in the estimator s^2?

Problem 4: A researcher estimates the following quadratic regression using data on $n = 420$ school districts:

$$\widehat{TestScore} = \underset{(2.9)}{607.3} + \underset{(0.27)}{3.85} \cdot Income - \underset{(0.0048)}{0.0423} \cdot Income^2, \quad R^2 = 0.554,$$

where $Income$ is average district income in thousands of dollars and standard errors are in parentheses.

a. Compute the predicted change in test scores when income increases from 20 to 21 (i.e., from \$20,000 to \$21,000).

b. Write $\Delta\hat{Y}$ from part (a) as a linear combination of $\hat{\beta}_1$ and $\hat{\beta}_2$. Using the general variance formula for a linear combination $a\hat{\beta}_1 + b\hat{\beta}_2$, write down the expression for $\text{Var}(\Delta\hat{Y})$ in terms of $\text{Var}(\hat{\beta}_1)$, $\text{Var}(\hat{\beta}_2)$, and $\text{Cov}(\hat{\beta}_1, \hat{\beta}_2)$.

c. Test the null hypothesis $H_0 : \beta_2 = 0$ against $H_1 : \beta_2 \neq 0$ at the 5% significance level. What does rejection of this hypothesis imply about the relationship between income and test scores?

d. At what value of $Income$ does the estimated regression function reach its maximum? Is this value within the sample range (5.3 to 55.3)? What does this tell you about the practical relevance of the quadratic specification?

Problem 5: Consider the following three estimated regressions, all using the same data on earnings (Y), years of education (X_1), and a binary variable D equal to 1 for females:

Regression A (Log-linear):

$$\widehat{\ln(Y)} = 2.10 + 0.085\, X_1$$

Regression B (Log-log):

$$\widehat{\ln(Y)} = 0.75 + 1.05\ \ln(X_1)$$

Regression C (Interaction):

$$\widehat{\ln(Y)} = 2.30 + 0.092\, X_1 - 0.40\, D - 0.015\, (X_1 \times D)$$

a. In Regression A, interpret the coefficient 0.085 in economic terms. What is the predicted percentage change in earnings from one additional year of education?

b. In Regression B, interpret the coefficient 1.05. What type of economic concept does β_1 represent in a log-log specification?

c. Can you directly compare the R^2 of Regression A with the R^2 of a linear regression $\hat{Y} = b_0 + b_1 X_1$? Why or why not?

d. In Regression C, compute the predicted effect of one additional year of education on $\ln(Y)$ for males ($D = 0$) and for females ($D = 1$). What is the economic interpretation of the coefficient -0.015 on the interaction term?

e. Using Regression C, a researcher claims: "The gender earnings gap is -0.40, or about 40%." Under what condition on X_1 is this statement correct? Write down the general expression for the gender gap as a function of X_1.

Note: In addition to the theoretical problems above, we will work through two empirical exercises during the recitation: (1) comparing **FGLS vs. OLS** estimation, and (2) **testing for heteroskedasticity**. The corresponding datasets and notebooks will be provided in class.