

**International School of Economics at TSU**  
**Econometrics 2**  
**Lasha Chochua**

**Problem Set 6 – Solutions**

**Problem 1: OLS Variance Under Heteroskedasticity**

a. We need to show  $\mathbb{E}[\hat{\beta} | \mathbf{X}] = \beta$ .

Substitute  $\mathbf{Y} = \mathbf{X}\beta + \mathbf{e}$  into the OLS formula:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta + \mathbf{e}) = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e}.$$

Taking the conditional expectation:

$$\mathbb{E}[\hat{\beta} | \mathbf{X}] = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \mathbb{E}[\mathbf{e} | \mathbf{X}] = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \mathbf{0} = \beta.$$

The key point: unbiasedness only requires  $\mathbb{E}[\mathbf{e} | \mathbf{X}] = \mathbf{0}$ . It does **not** require any assumption about  $\text{Var}[\mathbf{e} | \mathbf{X}]$ . Therefore OLS remains unbiased regardless of heteroskedasticity. ■

b. From part (a), we have  $\hat{\beta} - \beta = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e}$ .

The conditional variance is:

$$\text{Var}[\hat{\beta} | \mathbf{X}] = \mathbb{E} [(\hat{\beta} - \beta)(\hat{\beta} - \beta)' | \mathbf{X}].$$

Substituting:

$$= \mathbb{E} [(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e}\mathbf{e}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} | \mathbf{X}].$$

Since  $(\mathbf{X}'\mathbf{X})^{-1}$  and  $\mathbf{X}'$  are non-random conditional on  $\mathbf{X}$ , they pass through the expectation:

$$= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \mathbb{E}[\mathbf{e}\mathbf{e}' | \mathbf{X}] \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}.$$

Using  $\mathbb{E}[\mathbf{e}\mathbf{e}' | \mathbf{X}] = \text{Var}[\mathbf{e} | \mathbf{X}] = \sigma^2\Sigma$ :

$$\boxed{\text{Var}[\hat{\beta} | \mathbf{X}] = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\Sigma\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1}.} \quad \blacksquare$$

c. The standard estimator  $\widehat{V}_{\hat{\beta}}^0 = s^2(\mathbf{X}'\mathbf{X})^{-1}$  is derived under the assumption that  $\text{Var}[\mathbf{e} \mid \mathbf{X}] = \sigma^2 I_n$ , which gives  $\text{Var}[\widehat{\beta} \mid \mathbf{X}] = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ . This formula is correct only when the sandwich middle collapses:

$$\mathbf{X}'\Sigma\mathbf{X} = \mathbf{X}'I_n\mathbf{X} = \mathbf{X}'\mathbf{X}.$$

When  $\Sigma \neq I_n$ , the true variance has the sandwich form from part (b), which generally differs from  $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ . Therefore  $s^2(\mathbf{X}'\mathbf{X})^{-1}$  targets the wrong quantity – it is a biased estimator of the true variance.

The bias vanishes if and only if  $\Sigma = c \cdot I_n$  for some scalar  $c > 0$  (i.e., homoskedasticity up to a common scaling factor, which is absorbed into  $\sigma^2$ ). More precisely, the condition is  $\mathbf{X}'\Sigma\mathbf{X} = c \cdot \mathbf{X}'\mathbf{X}$ , which holds whenever  $\Sigma$  is proportional to the identity.

d.

- **Unbiasedness:** No, heteroskedasticity does **not** affect unbiasedness. As shown in part (a),  $\mathbb{E}[\widehat{\beta} \mid \mathbf{X}] = \beta$  requires only  $\mathbb{E}[\mathbf{e} \mid \mathbf{X}] = \mathbf{0}$ , which has nothing to do with the variance structure.
- **Efficiency:** Yes, heteroskedasticity **does** affect efficiency. Under homoskedasticity, the Gauss–Markov theorem guarantees that OLS is BLUE (Best Linear Unbiased Estimator). When  $\Sigma \neq I_n$ , OLS is still unbiased and linear, but it is no longer the most efficient linear unbiased estimator. The GLS estimator (Problem 2) achieves a lower variance – by Aitken’s theorem,  $\text{Var}[\tilde{\beta}_{glS}] \leq \text{Var}[\widehat{\beta}]$  in the matrix (positive semi-definite) sense.

## Problem 2: GLS as Transformed OLS

a. We compute the variance of the transformed errors  $\tilde{\mathbf{e}} = \Sigma^{-1/2}\mathbf{e}$ :

$$\text{Var}[\tilde{\mathbf{e}} \mid \mathbf{X}] = \Sigma^{-1/2} \text{Var}[\mathbf{e} \mid \mathbf{X}] (\Sigma^{-1/2})'.$$

Since  $\Sigma^{-1/2}$  is symmetric,  $(\Sigma^{-1/2})' = \Sigma^{-1/2}$ . Substituting  $\text{Var}[\mathbf{e} \mid \mathbf{X}] = \sigma^2\Sigma$ :

$$\text{Var}[\tilde{\mathbf{e}} \mid \mathbf{X}] = \Sigma^{-1/2}(\sigma^2\Sigma)\Sigma^{-1/2} = \sigma^2\Sigma^{-1/2}\Sigma\Sigma^{-1/2} = \sigma^2 I_n. \quad \blacksquare$$

This is the key insight: the transformation “whitens” the errors, restoring homoskedasticity.

b. The OLS estimator applied to the transformed model  $\tilde{\mathbf{Y}} = \tilde{\mathbf{X}}\beta + \tilde{\mathbf{e}}$  is:

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

Substituting  $\tilde{\mathbf{X}} = \Sigma^{-1/2}\mathbf{X}$  and  $\tilde{\mathbf{Y}} = \Sigma^{-1/2}\mathbf{Y}$ :

$$\tilde{\mathbf{X}}'\tilde{\mathbf{X}} = (\Sigma^{-1/2}\mathbf{X})'(\Sigma^{-1/2}\mathbf{X}) = \mathbf{X}'(\Sigma^{-1/2})'\Sigma^{-1/2}\mathbf{X} = \mathbf{X}'\Sigma^{-1}\mathbf{X},$$

where we used  $(\Sigma^{-1/2})'\Sigma^{-1/2} = \Sigma^{-1/2}\Sigma^{-1/2} = \Sigma^{-1}$  (since  $\Sigma^{-1/2}$  is symmetric).

Similarly:

$$\tilde{\mathbf{X}}'\tilde{\mathbf{Y}} = \mathbf{X}'\Sigma^{-1/2}\Sigma^{-1/2}\mathbf{Y} = \mathbf{X}'\Sigma^{-1}\mathbf{Y}.$$

Combining:

$$\boxed{\tilde{\beta}_{gl_s} = (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}\mathbf{Y}. \quad \blacksquare}$$

c. Since the transformed model satisfies the classical assumptions (from part a), we can apply the standard OLS variance formula to the transformed model:

$$\text{Var}[\tilde{\beta}_{gl_s} | \mathbf{X}] = \sigma^2(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}.$$

From part (b),  $\tilde{\mathbf{X}}'\tilde{\mathbf{X}} = \mathbf{X}'\Sigma^{-1}\mathbf{X}$ , so:

$$\boxed{\text{Var}[\tilde{\beta}_{gl_s} | \mathbf{X}] = \sigma^2(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}. \quad \blacksquare}$$

Note the simplicity compared to the OLS sandwich form in Problem 1(b). This is a single inverse rather than a triple product – a direct consequence of the transformed errors being homoskedastic.

d. When  $\Sigma = I_n$ :

- $\Sigma^{-1} = I_n$ , so  $\mathbf{X}'\Sigma^{-1}\mathbf{X} = \mathbf{X}'I_n\mathbf{X} = \mathbf{X}'\mathbf{X}$ .
- $\mathbf{X}'\Sigma^{-1}\mathbf{Y} = \mathbf{X}'I_n\mathbf{Y} = \mathbf{X}'\mathbf{Y}$ .

Substituting into the GLS formula:

$$\tilde{\beta}_{gl_s} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \hat{\beta}_{OLS}. \quad \blacksquare$$

This confirms that OLS is a special case of GLS when the errors are homoskedastic.

### Problem 3: HC0 vs. Classical Standard Errors

a. The sandwich form is:

$$\text{Var}[\hat{\beta} | \mathbf{X}] = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'D\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1},$$

where  $D = \text{diag}(\sigma_1^2, \dots, \sigma_n^2)$ . For the simple regression with  $\mathbf{X}_i = (1, X_i)'$ , the middle matrix is:

$$\mathbf{X}'D\mathbf{X} = \sum_{i=1}^n \sigma_i^2 \mathbf{X}_i \mathbf{X}_i' = \sum_{i=1}^n \sigma_i^2 \begin{bmatrix} 1 \\ X_i \end{bmatrix} \begin{bmatrix} 1 & X_i \end{bmatrix} = \begin{bmatrix} \sum \sigma_i^2 & \sum \sigma_i^2 X_i \\ \sum \sigma_i^2 X_i & \sum \sigma_i^2 X_i^2 \end{bmatrix}.$$

Under homoskedasticity ( $\sigma_i^2 = \sigma^2$  for all  $i$ ), this simplifies to  $\sigma^2 \mathbf{X}'\mathbf{X}$  and the sandwich collapses to  $\sigma^2 (\mathbf{X}'\mathbf{X})^{-1}$ . Under heteroskedasticity, it does not.

**b.** The classical standard error assumes  $\text{Var}[\hat{\beta} \mid \mathbf{X}] = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}$ , which requires  $\mathbf{X}'D\mathbf{X} = \sigma^2 \mathbf{X}'\mathbf{X}$ .

When  $\sigma_i^2 = \sigma^2 X_i^2$ , the (2, 2) entry of the middle matrix becomes  $\sum \sigma^2 X_i^2 \cdot X_i^2 = \sigma^2 \sum X_i^4$ , whereas the (2, 2) entry of  $\sigma^2 \mathbf{X}'\mathbf{X}$  is  $\sigma^2 \sum X_i^2$ . These are generally different (they coincide only if all  $X_i^2 = 1$ , which is a degenerate case). Since the middle of the sandwich is misspecified, the classical standard error computes the wrong quantity – it could be too large or too small depending on the data, leading to incorrect  $t$ -statistics and confidence intervals.

**c.** The key insight comes from what we already know about  $s^2$ . Recall that under homoskedasticity:

$$\mathbb{E} \left[ \sum_{i=1}^n \hat{e}_i^2 \mid \mathbf{X} \right] = (n - k) \sigma^2, \quad \text{not } n \sigma^2.$$

This tells us that **on average**, the sum of squared residuals is smaller than the sum of squared errors. Why? OLS chooses  $\hat{\beta}$  to **minimize**  $\sum \hat{e}_i^2$ . The fitted values  $\hat{Y}_i = \mathbf{X}_i' \hat{\beta}$  are chosen to be as close to  $Y_i$  as possible, so the residuals  $\hat{e}_i = Y_i - \hat{Y}_i$  are systematically “shrunk” toward zero compared to the true errors  $e_i = Y_i - \mathbf{X}_i' \beta$ .

Estimating  $k$  parameters uses up  $k$  degrees of freedom, which is why the total  $\sum \hat{e}_i^2$  underestimates  $\sum e_i^2$  by a factor of  $\frac{n-k}{n}$  on average. This means that at the individual level,  $\hat{e}_i^2$  tends to underestimate  $\sigma_i^2 = \mathbb{E}[e_i^2 \mid \mathbf{X}]$ .

The bias is **downward**:  $\mathbb{E}[\hat{e}_i^2 \mid \mathbf{X}] \leq \sigma_i^2$  on average across observations. This is the same phenomenon that makes  $\frac{1}{n} \sum \hat{e}_i^2$  a biased estimator of  $\sigma^2$  – the residuals understate the true error dispersion because OLS has already “used” some of the variation in  $\mathbf{Y}$  to fit the coefficients.

**d.** The HC1 correction multiplies HC0 by  $\frac{n}{n-k}$ :

$$\hat{V}_{\hat{\beta}}^{\text{HC1}} = \frac{n}{n-k} \hat{V}_{\hat{\beta}}^{\text{HC0}}.$$

The intuition is identical to the degrees-of-freedom correction in  $s^2 = \frac{1}{n-k} \sum \hat{e}_i^2$  versus  $\hat{\sigma}^2 = \frac{1}{n} \sum \hat{e}_i^2$ . The OLS residuals satisfy  $k$  orthogonality conditions ( $\mathbf{X}'\hat{\mathbf{e}} = \mathbf{0}$ ), so only  $n - k$  of the  $n$  residuals are “free.” Dividing by  $n$  (or equivalently, not scaling HC0) systematically underestimates the variance. The factor  $\frac{n}{n-k} > 1$  inflates the estimate to compensate.

Note that HC1 is an approximate correction – it does not make the estimator exactly unbiased observation-by-observation, but it improves small-sample performance in practice.

#### Problem 4: Interpreting Nonlinear Regression Functions

a. Using the estimated quadratic regression:

$$\widehat{TestScore} = 607.3 + 3.85 \cdot Income - 0.0423 \cdot Income^2.$$

Predicted value at  $Income = 20$ :

$$\hat{Y}_{20} = 607.3 + 3.85(20) - 0.0423(20^2) = 607.3 + 77.0 - 16.92 = 667.38.$$

Predicted value at  $Income = 21$ :

$$\hat{Y}_{21} = 607.3 + 3.85(21) - 0.0423(21^2) = 607.3 + 80.85 - 18.65 = 669.50.$$

Therefore:

$$\Delta \hat{Y} = 669.50 - 667.38 = 2.12 \text{ points.}$$

b. We can write:

$$\Delta \hat{Y} = \hat{\beta}_1(21 - 20) + \hat{\beta}_2(21^2 - 20^2) = \hat{\beta}_1(1) + \hat{\beta}_2(441 - 400) = \hat{\beta}_1 + 41\hat{\beta}_2.$$

This is a linear combination  $a\hat{\beta}_1 + b\hat{\beta}_2$  with  $a = 1$  and  $b = 41$ .

Applying the variance formula  $\text{Var}(a\hat{\beta}_1 + b\hat{\beta}_2) = a^2\text{Var}(\hat{\beta}_1) + b^2\text{Var}(\hat{\beta}_2) + 2ab\text{Cov}(\hat{\beta}_1, \hat{\beta}_2)$ :

$$\text{Var}(\Delta \hat{Y}) = \text{Var}(\hat{\beta}_1) + 41^2 \text{Var}(\hat{\beta}_2) + 2(41) \text{Cov}(\hat{\beta}_1, \hat{\beta}_2).$$

The standard error is  $SE(\Delta \hat{Y}) = \sqrt{\text{Var}(\Delta \hat{Y})}$ , and a 95% confidence interval is  $\Delta \hat{Y} \pm 1.96 \cdot SE(\Delta \hat{Y})$ .

c. The  $t$ -statistic for  $H_0 : \beta_2 = 0$  is:

$$t = \frac{\hat{\beta}_2}{SE(\hat{\beta}_2)} = \frac{-0.0423}{0.0048} = -8.81.$$

The critical value for a two-sided test at the 5% level (with  $n - 3 = 417$  degrees of freedom, approximately  $z$ ) is  $\pm 1.96$ . Since  $|t| = 8.81 > 1.96$ , we **reject**  $H_0$  at the 5% level (indeed, the  $p$ -value is less than 0.01%).

**Implication:** The relationship between income and test scores is **not linear**. The statistically significant negative coefficient on  $Income^2$  means that the marginal effect of income on test scores **diminishes** as income increases – the relationship exhibits decreasing returns.

d. The quadratic  $f(x) = 607.3 + 3.85x - 0.0423x^2$  reaches its maximum where the first derivative equals zero:

$$f'(x) = 3.85 - 2(0.0423)x = 3.85 - 0.0846x = 0 \implies x^* = \frac{3.85}{0.0846} \approx 45.5.$$

This value is **within** the sample range of [5.3, 55.3]. This means the estimated regression function predicts that test scores **increase** with income up to about \$45,500, and then **decrease** for higher incomes.

Whether this turning point is economically meaningful requires caution. There are very few observations at income levels above \$45,000, so the downward-sloping portion of the quadratic is driven by limited data. It may be a mechanical artifact of the quadratic functional form rather than a genuine decrease in test scores at high incomes. A linear-log specification, which has no turning point, might be more appropriate for this reason.

### Problem 5: Logarithmic Specifications and Interaction Terms

a. Regression A is a **log-linear** model:  $\ln(Y) = \beta_0 + \beta_1 X_1 + u$ .

In this specification, a one-unit change in  $X_1$  (one additional year of education) is associated with a  $100 \times \beta_1 = 100 \times 0.085 = 8.5\%$  change in earnings.

Interpretation: Each additional year of education is associated with an approximate **8.5% increase** in earnings, holding everything else constant.

b. Regression B is a **log-log** model:  $\ln(Y) = \beta_0 + \beta_1 \ln(X_1) + u$ .

In this specification,  $\beta_1$  is the **elasticity** of  $Y$  with respect to  $X_1$ : a 1% increase in years of education is associated with a  $\beta_1 = 1.05\%$  increase in earnings.

Since the elasticity is approximately 1, this means earnings increase roughly proportionally with education. Precisely, a 1% increase in education years is associated with a 1.05% increase in earnings.

c. No, the  $R^2$  values **cannot** be directly compared. The  $R^2$  measures the fraction of the variance of the **dependent variable** explained by the regressors. In Regression A, the dependent variable is  $\ln(Y)$ , whereas in the linear regression, it is  $Y$  itself.

Since  $\text{Var}[\ln(Y)] \neq \text{Var}[Y]$ , the denominators of the two  $R^2$  formulas are different, making the values non-comparable. An  $R^2$  of 0.50 for  $\ln(Y)$  does not mean the same thing as an  $R^2$  of 0.50 for  $Y$ .

To compare these models, one would need to either: (i) transform the predictions back to the original scale and compute a comparable measure, or (ii) use economic reasoning to decide whether  $Y$  or  $\ln(Y)$  is the more natural dependent variable.

d. Regression C is:  $\widehat{\ln(Y)} = 2.30 + 0.092 X_1 - 0.40 D - 0.015 (X_1 \times D)$ .

The effect of one additional year of education on  $\ln(Y)$  is found by computing  $\frac{\Delta \widehat{\ln(Y)}}{\Delta X_1}$ :

- For **males** ( $D = 0$ ): effect =  $0.092 + (-0.015)(0) = 0.092$ , i.e., approximately **9.2%** increase in earnings per additional year.
- For **females** ( $D = 1$ ): effect =  $0.092 + (-0.015)(1) = 0.077$ , i.e., approximately **7.7%** increase in earnings per additional year.

The coefficient  $-0.015$  on the interaction term is the **difference** in the return to education between females and males. It means that each additional year of education increases female earnings by 1.5 percentage points **less** than it increases male earnings. In other words, the return to education is lower for women than for men.

e. The gender earnings gap is the difference in  $\widehat{\ln(Y)}$  between females ( $D = 1$ ) and males ( $D = 0$ ), holding  $X_1$  fixed:

$$\begin{aligned} \widehat{\ln(Y)}|_{D=1} - \widehat{\ln(Y)}|_{D=0} &= (2.30 + 0.092 X_1 - 0.40 - 0.015 X_1) - (2.30 + 0.092 X_1) \\ &= -0.40 - 0.015 X_1. \end{aligned}$$

So the gender gap depends on education:

$$\boxed{\text{Gender gap} = -0.40 - 0.015 X_1.}$$

The researcher's claim that the gap is " $-0.40$ , or about 40%" is correct **only when**  $X_1 = 0$  (zero years of education). For any positive level of education, the gap is **larger** than 40% (more negative). For example:

- At  $X_1 = 12$  years: gap =  $-0.40 - 0.015(12) = -0.58$ , i.e., approximately 58%.
- At  $X_1 = 16$  years: gap =  $-0.40 - 0.015(16) = -0.64$ , i.e., approximately 64%.

Thus the interaction term implies that the gender earnings gap **widens** with education – women with more education face a proportionally larger gap relative to men with the same education level.