Liming Lin
Professor Moshe Buchinsky
Econometrics II
Problem Set 5
Apr. 1st, 2025

Problem Set 5

Problem 1

Consider the simple regression $y = \alpha_\beta x + u$. By the Central Limit Theorem, we know that $\sum_{i=1}^{N} \frac{u_i}{N} \xrightarrow{d} N(0, \frac{\sigma^2}{N})$. Show that: $\sqrt{N}(\hat{\beta} - \beta) \xrightarrow{d} N\left(0, \frac{\sigma^2}{\text{Var}(x_i)}\right)$

Proof: We are given the simple linear regression model:

$$y_i = \alpha + \beta x_i + u_i$$

with $\mathbb{E}[u_i|x_i] = 0$ and $\text{Var}(u_i|x_i) = \sigma^2$.
The OLS estimator for β is given by:

$$\hat{\beta} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

We substitute the model $y_i = \alpha + \beta x_i + u_i$ into the OLS formula. First note:

$$\bar{y} = \alpha + \beta\bar{x} + \bar{u}$$

Since $\bar{u} = \frac{1}{N}\sum_{i=1}^{N} u_i \to 0$ as N grows large (and is exactly zero in finite samples with an intercept), we approximate:

$$y_i - \bar{y} = \beta(x_i - \bar{x}) + (u_i - \bar{u}) = \beta(x_i - \bar{x}) + u_i$$

Thus:

$$\begin{aligned}\hat{\beta} &= \frac{\sum_{i=1}^{N}(x_i - \bar{x})\left[\beta(x_i - \bar{x}) + u_i\right]}{\sum_{i=1}^{N}(x_i - \bar{x})^2} \\ &= \beta \cdot \frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2} + \frac{\sum_{i=1}^{N}(x_i - \bar{x})u_i}{\sum_{i=1}^{N}(x_i - \bar{x})^2} \\ &= \beta + \frac{\sum_{i=1}^{N}(x_i - \bar{x})u_i}{\sum_{i=1}^{N}(x_i - \bar{x})^2}\end{aligned}$$

So,

$$\hat{\beta} - \beta = \frac{\sum_{i=1}^{N}(x_i - \bar{x})u_i}{\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

Multiplying both sides by $\sqrt{N}$:

$$\sqrt{N}(\hat{\beta} - \beta) = \frac{\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(x_i - \bar{x})u_i}{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

Define:

$$A_N = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(x_i - \bar{x})u_i$$

$$B_N = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2$$

- By the Central Limit Theorem:

$$A_N \xrightarrow{d} N\left(0, \sigma^2 \cdot \text{Var}(x_i)\right)$$

- By the Law of Large Numbers:

$$B_N \xrightarrow{p} \text{Var}(x_i)$$

Slutsky's Theorem states that if $A_N \xrightarrow{d} A$ and $B_N \xrightarrow{p} B \neq 0$, then:

$$\frac{A_N}{B_N} \xrightarrow{d} \frac{A}{B}$$

We apply it here:

$$\sqrt{N}(\hat{\beta} - \beta) = \frac{A_N}{B_N} \xrightarrow{d} \frac{N\left(0, \sigma^2 \cdot \text{Var}(x_i)\right)}{\text{Var}(x_i)} = N\left(0, \frac{1}{(Var(x_i))^2}\sigma^2 Var(x_i)\right) = N\left(0, \frac{\sigma^2}{\text{Var}(x_i)}\right)$$

We have shown, using substitution, the Central Limit Theorem, the Law of Large Numbers, and Slutsky's Theorem, that:

$$\sqrt{N}(\hat{\beta} - \beta) \xrightarrow{d} N\left(0, \frac{\sigma^2}{\text{Var}(x_i)}\right)$$

as required.

Problem 2

Consider the simple linear regression model without constant:

$$y_i = \beta x_i + \epsilon_i, i = 1, ..., n$$

Suppose that the homoskedasticity assumption doesn't hold. Instead the conditional variance of the error term is assumed to vary with the regressor: $\mathbb{V}(\epsilon_i | x_i) = \sigma^2$

1. Find the OLS estimator $\hat{\beta}$ and determine its conditional variance $\mathbb{V}(\hat{\beta} | x_1, ..., x_n)$.

   Answers: The OLS estimator minimizes the sum of squared residuals:

   $$\sum_{i=1}^{n} (y_i - \beta x_i)^2$$

   Taking the derivative w.r.t. β and setting it to zero gives:

   $$\hat{\beta} = \frac{\sum x_i y_i}{\sum x_i^2}$$

   Substituting the model $y_i = \beta x_i + \varepsilon_i$ into the estimator and use similar methods as in Problem 1:

   $$\hat{\beta} = \frac{\sum x_i(\beta x_i + \varepsilon_i)}{\sum x_i^2} = \beta + \frac{\sum x_i \varepsilon_i}{\sum x_i^2}$$

   To compute the conditional variance, we treat x_i as fixed (conditioning on $x_1, \ldots, x_n$). Then:

   $$\text{Var}(\hat{\beta} \mid x) = \text{Var}\left(\frac{\sum x_i \varepsilon_i}{\sum x_i^2}\right) = \frac{1}{\left(\sum x_i^2\right)^2} \sum x_i^2 \cdot \text{Var}(\varepsilon_i \mid x_i) = \frac{\sum x_i^2 \sigma_i^2}{\left(\sum x_i^2\right)^2}$$

2. Suppose that $\sigma_i^2 = \sigma^2 x_i$ (x_i is assumed positive for all i), and consider the transformed model:

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

where $y_i^* = \frac{y_i}{\sqrt{x_i}}, \quad x_i^* = \frac{x_i}{\sqrt{x_i}}, \quad \varepsilon_i^* = \frac{\varepsilon_i}{\sqrt{x_i}}$

Find the GLS estimator $\hat{\beta}^*$ and its conditional variance $\mathbb{V}(\hat{\beta}^* | x_1, ..., x_n)$.

Answers:
Note that $x_i^* = \frac{x_i}{\sqrt{x_i}} = \sqrt{x_i}$,

We compute the variance of the transformed errors:

$$\text{Var}(\varepsilon_i^* \mid x_i) = \frac{\text{Var}(\varepsilon_i \mid x_i)}{(\sqrt{x_i})^2} = \frac{\sigma^2 x_i}{x_i} = \sigma^2$$

So the transformed model has homoskedastic errors.

Apply OLS to the transformed model:

$$\beta^* = \frac{\sum x_i^* y_i^*}{\sum (x_i^*)^2} = \frac{\sum \sqrt{x_i} \cdot \frac{y_i}{\sqrt{x_i}}}{\sum (\sqrt{x_i})^2} = \frac{\sum y_i}{\sum x_i}$$

Conditional Variance: under the normal assumption, the variance of the OLS estimator is:

$$Var(\hat{\beta} \mid x) = \frac{\sigma^2}{\sum (x_i)^2}$$

then we plug in the transformed model:

$$\text{Var}(\beta^* \mid x) = \frac{\sigma^2}{\sum (\sqrt{x_i})^2} = \frac{\sigma^2}{\sum x_i}$$

3. Show that $\mathbb{V}(\beta^* \mid x_1^*, ..., x_n^*) \leq \mathbb{V}(\hat{\beta} \mid x_1, ..., x_n)$

Proof:
Plug in $\sigma_i^2 = \sigma^2 x_i$

$$\text{Var}(\hat{\beta}) = \sigma^2 \cdot \frac{\sum x_i^3}{\left(\sum x_i^2\right)^2}, \quad \text{Var}(\beta^*) = \frac{\sigma^2}{\sum x_i}$$

Dividing both sides by σ^2, we need to show:

$$\frac{1}{\sum x_i} \leq \frac{\sum x_i^3}{\left(\sum x_i^2\right)^2}$$

Set $a_i = x_i^{1.5}$ and $b_i = x_i^{0.5}$:

$$\frac{1}{\sum b_i^2} \leq \frac{\sum a_i^2}{\sum (a_i b_i)^2}$$

Reaarranging gives:

$$\left(\sum a_i b_i\right)^2 \leq \sum a_i^2 \cdot \sum b_i^2$$

This inequality holds due to the Cauchy–Schwarz inequality. Therefore:

$$\text{Var}(\beta^*) \leq \text{Var}(\hat{\beta})$$

4. Someone ignoring the heteroskedasticity problem might (wrongly) assume that the conditional variance of the OLS estimator is $\sigma^2 / \sum x_i^2$, and estimate this variance by

$$\frac{1}{n-1} \frac{\sum (y_i - \hat{\beta} x_i)^2}{\sum x_i^2}$$

Calculate the conditional expectation of the above estimator, and show that it is smaller than the true varaince of the OLS estimator (obtain in question 1). What does that imply for statistical inference (confidence intervals, Type I error)?

Answers:
First, expand the squared residual:

$$y_i - \hat{\beta} x_i = \beta x_i + \epsilon_i - \hat{\beta} x_i = \epsilon_i - (\hat{\beta} - \beta) x_i$$

$$(y_i - \hat{\beta} x_i)^2 = \epsilon_i^2 - 2(\hat{\beta} - \beta) x_i \epsilon_i + (\hat{\beta} - \beta)^2 x_i^2$$

Take the expectation conditional on x:

$$\mathbb{E}[\epsilon_i^2 \mid x_i] = \sigma_i^2$$

$$\mathbb{E}[(\hat{\beta} - \beta) x_i \epsilon_i] = (\hat{\beta} - \beta) x_i \cdot 0$$

because the heteroskedasticity problem is ignored.

$$\mathbb{E}[(\hat{\beta} - \beta)^2 x_i^2] = \mathrm{Var}(\hat{\beta}) \cdot x_i^2$$

Summing the conditional expectations:

$$\mathbb{E}[\sum (y_i - \hat{\beta} x_i)^2 \mid x] = \sum \sigma_i^2 + \mathrm{Var}(\hat{\beta}) \cdot \sum x_i^2$$

Then:

$$\mathbb{E}[\widehat{\mathrm{Var}}(\hat{\beta}) \mid x] = \frac{1}{n-1} \cdot \left(\frac{\sum \sigma_i^2}{\sum x_i^2} + \mathrm{Var}(\hat{\beta}) \right)$$

We want to show that :

$$\mathbb{E}[\widehat{\mathrm{Var}}(\hat{\beta}) \mid x] < \mathrm{Var}(\hat{\beta})$$

Substituting the expressionfrom previous questions:

$$\frac{1}{n-1} \cdot \left(\frac{\sum \sigma_i^2}{\sum x_i^2} + \mathrm{Var}(\hat{\beta}) \right) < \frac{\sum x_i^2 \sigma_i^2}{\left(\sum x_i^2 \right)^2}$$

Multiplying both sides by $(n-1)$ and substracting $\mathrm{Var}(\hat{\beta})$ gives:

$$\frac{\sum \sigma_i^2}{\sum x_i^2} < \frac{\sum x_i^2 \sigma_i^2}{\left(\sum x_i^2 \right)^2} \cdot (n-1) - \mathrm{Var}(\hat{\beta})$$

$$\frac{\sum \sigma_i^2}{\sum x_i^2} < \frac{\sum x_i^2 \sigma_i^2}{\left(\sum x_i^2 \right)^2} \cdot (n-2)$$

This inequality holds because the left-hand side is the average of the variances, while the right-hand side is a weighted average of the variances divided by $n-2$.

Implications for Inference

- Estimated standard errors are too small
- t-statistics are too large
- Confidence intervals are too narrow
- Type I error increases — more false rejections of true null hypotheses
- Inference becomes unreliable