EC 306: Introductory Econometrics

Class Problem Sheet 2

Hints/Sketches of answers are provided in italics below the questions.

1. The simple linear regression model without intercept was discussed in the lectures and is given by:

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

where X_i is a scalar. For this question, we will free up one of the classical assumptions to allow for heteroskedasticity. In particular, the classical assumptions hold except we now assume

$$var(\varepsilon_i) = \sigma^2 \omega_i^2.$$

i) The ordinary least squares estimator is given by:

$$\widehat{\beta} = \frac{\sum X_i Y_i}{\sum X_i^2}.$$

Show that $E\left(\widehat{\beta}\right) = \beta$ and, thus, that the OLS estimator is unbiased. What is $var\left(\widehat{\beta}\right)$?

SKETCH OF ANSWER: The OLS estimator can be written as:

$$\widehat{\beta} = \beta + \frac{\sum X_i \varepsilon_i}{\sum X_i^2}$$

Taking the expected value of both sides of this equation and using the properties of the expectations operator, the fact that $E\left(\widehat{\beta}\right) = \beta$ can be established in a similar fashion as in some of the lecture proofs. The key step in the proof involves noting that X_i is assumed to be fixed (not a random variable) and hence we end up with a form involving $E(\varepsilon_i)$ and can use the fact that the errors have mean zero to establish unbiasedness.

We can derive the variance using various properties of variance operator:

$$
\begin{aligned}
var\left(\widehat{\beta}\right) &= var\left(\beta + \frac{\sum X_i \varepsilon_i}{\sum X_i^2}\right) \\
&= var\left(\frac{\sum X_i \varepsilon_i}{\sum X_i^2}\right) \\
&= \frac{1}{\left(\sum X_i^2\right)^2} var\left(\sum X_i \varepsilon_i\right) \\
&= \frac{1}{\left(\sum X_i^2\right)^2} \sum X_i^2 var\left(\varepsilon_i\right) \\
&= \frac{\sigma^2}{\left(\sum X_i^2\right)^2} \sum X_i^2 \omega_i^2
\end{aligned}
$$

ii) The generalized least squares estimator for this model is given by:

$$
\widehat{\beta}_{GLS} = \frac{\sum \left(\frac{X_i}{\omega_i}\right)\left(\frac{y_i}{\omega_i}\right)}{\sum \left(\frac{X_i}{\omega_i}\right)^2}
$$

Show that $E\left(\widehat{\beta}_{GLS}\right) = \beta$ and, thus, that the OLS estimator is unbiased. What is $var\left(\widehat{\beta}_{GLS}\right)$? Is $var\left(\widehat{\beta}_{GLS}\right) \leq var\left(\widehat{\beta}\right)$?

ANSWER: as noted in lectures, if we define $Y_i^* = \frac{Y_i}{\omega_i}$ and $X_i^* = \frac{X_i}{\omega_i}$, *then the GLS estimator is simply an OLS estimator – only one which uses* Y_i^* *and* X_i^* *instead of* Y_i *and* X_i. *In other words, GLS is OLS on a transformed model. The key question is: "does this transformed model satisfy the classical assumptions?" The answer to this is "Yes". You should confirm this yourself, I will answer only one (the most crucial) part. Remember that the classical assumptions required* $var\left(Y_i\right) = \sigma^2$. *Under heteroskedasticity this does not hold and in this question we have* $var\left(Y_i\right) = \sigma^2 \omega_i^2$. *But*

$$
var\left(Y_i^*\right) = var\left(\frac{Y_i}{\omega_i}\right) = \frac{1}{\omega_i^2} var\left(Y_i\right) = \frac{\sigma^2 \omega_i^2}{\omega_i^2} = \sigma^2.
$$

Thus, the transformed model is homoskedastic.

Since the transformed model satisfies the classical assumptions and GLS is OLS on this transformed model, we can use all our old OLS results to derive everything asked for in this question. I will not repeat these derivations here. Note, however, that the Gauss Markov theorem implies immediately that $var\left(\widehat{\beta}_{GLS}\right) \leq var\left(\widehat{\beta}\right)$ *so just citing this theorem (and explaining why it is relevant here) is enough to prove the final part of the question.*

iii) Assume that you know what $var(\varepsilon_i)$ is and that $\widehat{\beta}_{GLS}$ is Normally distributed. Derive a 95% confidence interval involving the GLS estimator using your results form part ii).

ANSWER: .

The results derived in part ii), and the fact the linear combinations of Normal random variables are still Normal, imply that:

$$Z = \frac{\widehat{\beta}_{GLS} - \beta}{\sqrt{\frac{\sigma^2}{\sum \left(\frac{X_i}{\omega_i}\right)^2}}} \text{ is } N(0,1)$$

We can now use statistical tables for the Normal distribution to make probability statements:

$$\Pr\left[-1.96 \leq Z \leq 1.96\right] = 0.95$$

To get 95% confidence interval, we rearrange the inequalities to put β in the middle:

$$\Pr\left[-1.96 \leq \frac{\widehat{\beta}_{GLS} - \beta}{\sqrt{\frac{\sigma^2}{\sum \left(\frac{X_i}{\omega_i}\right)^2}}} \leq 1.96\right] = 0.95$$

rearranging:

$$\Pr\left[\widehat{\beta}_{GLS} - 1.96\sqrt{\frac{\sigma^2}{\sum \left(\frac{X_i}{\omega_i}\right)^2}} \leq \beta \leq \widehat{\beta}_{GLS} + 1.96\sqrt{\frac{\sigma^2}{\sum \left(\frac{X_i}{\omega_i}\right)^2}}\right] = 0.95$$

Thus, 95% confidence interval is

$$\left[\widehat{\beta}_{GLS} - 1.96\sqrt{\frac{\sigma^2}{\sum \left(\frac{X_i}{\omega_i}\right)^2}} \leq \beta \leq \widehat{\beta}_{GLS} + 1.96\sqrt{\frac{\sigma^2}{\sum \left(\frac{X_i}{\omega_i}\right)^2}}\right]$$

iv) Assume that you know what $\sigma^2\omega_i^2$ is and that $\widehat{\beta}$ is Normally distributed. Derive a 95% confidence interval involving the OLS estimator using your results from part i).

ANSWER: The derivation is the same as for part iii) except that the OLS estimator has a different variance:

$$Z = \frac{\widehat{\beta} - \beta}{\sqrt{\frac{\sigma^2}{\left(\sum X_i^2\right)^2} \sum X_i^2 \omega_i^2}} \text{ is } N(0,1)$$

We can now use statistical tables for the Normal distribution to make probability statements:

$$\Pr\left[-1.96 \leq Z \leq 1.96\right] = 0.95$$

To get 95% confidence interval, we rearrange the inequalities to put β *in the middle:*

$$\Pr\left[-1.96 \leq \frac{\widehat{\beta}-\beta}{\sqrt{\frac{\sigma^2}{\left(\sum X_i^2\right)^2}\sum X_i^2\omega_i^2}} \leq 1.96\right] = 0.95$$

rearranging:

$$\Pr\left[\widehat{\beta} - 1.96\sqrt{\frac{\sigma^2}{\left(\sum X_i^2\right)^2}\sum X_i^2\omega_i^2} \leq \beta \leq \widehat{\beta} + 1.96\sqrt{\frac{\sigma^2}{\left(\sum X_i^2\right)^2}\sum X_i^2\omega_i^2}\right] = 0.95$$

Thus, 95% confidence interval is

$$\left[\widehat{\beta} - 1.96\sqrt{\frac{\sigma^2}{\left(\sum X_i^2\right)^2}\sum X_i^2\omega_i^2} \leq \beta \leq \widehat{\beta} + 1.96\sqrt{\frac{\sigma^2}{\left(\sum X_i^2\right)^2}\sum X_i^2\omega_i^2}\right]$$

v) Now consider two possible 95% confidence intervals for β. They are the (correct) OLS and GLS confidence intervals given in parts iii) and iv). Compare these two intervals. Which one is wider?

ANSWER: The Gauss Markov theorem tells us that var $\left(\widehat{\beta}_{GLS}\right) \leq var\left(\widehat{\beta}\right)$ *(using the correct formula for var* $\left(\widehat{\beta}\right)$*). An examination of the manner in which these variances enter the confidence intervals shows that this means the GLS confidence interval is narrower than the OLS confidence interval (and, thus, that GLS is providing more accurate estimates). As a digression, note that you could also get a third confidence interval for OLS estimator which obtains under the classical assumptions (i.e. the standard one which was derived in Chapter 3 under the classical assumptions). If heteroskedasticity is present, this third confidence interval is incorrect and should never be used.*

2. (Measurement error in the dependent variable). Consider the regression model

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

This regression satisfies the classical assumptions. However, you cannot run this regression since do not observe Y_i, but instead observe:

$$Y_i^* = Y_i + v_i,$$

where v_i is i.i.d. with mean zero, variance σ_ν^2 and is independent of ε_i. Show that OLS is BLUE in the regression of Y^* on X.

ANSWER: If you put the expression for Y from the first equation into the second you get:

$$Y_i^* = \beta X_i + (\varepsilon_i + v_i)$$

This is a regression you can run in practice since the dependent variable and explanatory variable are now observed. The properties of OLS depend on what assumptions the regression satisfies. If it satisfies the classical assumptions, then we can call on the Gauss-Markov theorem to say that OLS is BLUE. Hence, to answer this question you must prove it satisfies the classical assumptions. I will not provide full details on the answer. One of the key assumptions is that the errors are not heteroskedastic. You can prove that this assumption holds as follows:

In this new regression, the error is $(\varepsilon_i + v_i)$. *You can use the properties of the variance operator (and the assumptions about* ε_i and v_i*) as follows:*

$$\begin{aligned} var\,(\varepsilon_i + v_i) &= var\,(\varepsilon_i) + var\,(v_i) \\ &= \sigma^2 + \sigma_v^2 \end{aligned}$$

This does not have exactly the same form as in Chapter 3, but that does not matter. The crucial thing is that the variance of the errors is constant. Therefore the second of the classical assumptions is satisfied. I leave you to prove that the other classical assumptions are also satisfied.

Remember that, in the lectures, we showed that if there was measurement error in the explanatory variable, then OLS was biased. The point of this exercise is to show that measurment error in the dependent variable does not cause any such problem.