Heteroskedasticity worksheet II

Ryan T. Godwin

GLS

1. What information does LS ignore that can lead to a more efficient estimator (in the presence of heteroskedasticity)?

   Observations with less variance should be given more weight than observations with high variance. The observations with smaller variance are more "valuable".

2. P is going to help us derive the GLS estimator. What is the relationship between Σ and P?

$$\Sigma^{-1} = P'P$$

3. When Σ is $diag(\sigma_1^2, \ldots, \sigma_n^2)$, what is the P matrix?

$$P = \begin{bmatrix} \frac{1}{\sigma_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\sigma_n} \end{bmatrix}$$

4. Write the formula for the GLS estimator. Then show how P can be used to transform the base population model to get the GLS estimator.

   The GLS estimator is:

$$\widehat{\boldsymbol{\beta}}_{\text{GLS}} = \left[X'\Sigma^{-1}X\right]^{-1} X'\Sigma^{-1}\boldsymbol{y}$$

   The P matrix can be used to transform the population model:

$$P\boldsymbol{y} = PX\boldsymbol{\beta} + P\boldsymbol{\epsilon}$$

   Applying LS to this model yields:

$$\begin{aligned} &\left[(PX)'(PX)\right]^{-1}(PX)'(P\boldsymbol{y}) \\ =&\left[X'P'PX\right]^{-1}X'P'P\boldsymbol{y} \\ =&\left[X'\Sigma^{-1}X\right]^{-1}X'\Sigma^{-1}\boldsymbol{y} \end{aligned}$$

   which is the GLS estimator.

Properties of GLS

5. Show that applying LS to the transformed model (using P) leads to an unbiased and consistent estimator.

   Since P is a non-random matrix, $E[P\boldsymbol{\epsilon}] = PE[\boldsymbol{\epsilon}]$, and assumption A.3 and A.5 are not affected. The LS estimator is still unbiased and consistent.

6. Show that the transformed model achieves A.4.

$$
\begin{aligned}
V\left[\boldsymbol{\epsilon}^*\right] &= V[P\boldsymbol{\epsilon}] \\
&= PV(\boldsymbol{\epsilon})P' \\
&= P(\Sigma)P' = P\Sigma P' \\
&= P\left(\Sigma^{-1}\right)^{-1}P' \\
&= P\left(P'P\right)^{-1}P' \\
&= PP^{-1}\left(P'\right)^{-1}P' = I
\end{aligned}
$$

7. Argue that GLS is more efficient than LS using the G-M theorem.

   The transformed model has an error-term that satisfies the usual assumptions. In particular, the transformed model is homoskedastic. So, if we apply LS to the transformed model, we'll get the BLU estimator of $\boldsymbol{\beta}$, by the Gauss-Markhov Theorem. This means that LS, under heteroskedasticity, is inefficient.

Unknown σ^2

8. Show that we just need to know Ω in order to do GLS.

   If we instead use $P'P = \Omega$ to get the P matrix, we get the same GLS estimator. That is, if:

$$
\Omega = \begin{bmatrix} \omega_{11} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \omega_{nn} \end{bmatrix}
$$

   then the P matrix used to transform the data can instead be written as:

$$
P = \begin{bmatrix} \omega_{11}^{-1/2} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \omega_{nn}^{-1/2} \end{bmatrix}
$$

   The LS estimator applied to the transformed model is:

$$
\begin{aligned}
&\left[(PX)'(PX)\right]^{-1}(PX)'(P\boldsymbol{y}) \\
=&\left[X'P'PX\right]^{-1}X'P'P\boldsymbol{y} \\
=&\left[X'\Omega^{-1}X\right]^{-1}X'\Omega^{-1}\boldsymbol{y} \\
=&\left[X'(\sigma^2\Omega)^{-1}X\right]^{-1}X'(\sigma^2\Omega)^{-1}\boldsymbol{y} \\
=&\left[X'\Sigma^{-1}X\right]^{-1}X'\Sigma^{-1}\boldsymbol{y}
\end{aligned}
$$

   which is the same GLS estimator derived before.

9. What is the interpretation of knowing Ω instead of Σ?

   We don't need to know the variance of each observation, we just need to know their proportionality.

10. Explain why GLS is also called "weighted least squares".

    GLS amounts to weighting the data by the inverse of the square root of the proportionality constants ω_{ii}:

$$\omega_{ii}^{-\frac{1}{2}} y_i = \beta_1 \omega_{ii}^{-\frac{1}{2}} + \beta_2 \left(\omega_{ii}^{-\frac{1}{2}} x_{i2} \right) + \cdots + \left(\omega_{ii}^{-\frac{1}{2}} \epsilon_i \right)$$

11. Suppose that $n = 3$. $\sigma_2^2 = 4\sigma_1^2$ and $\sigma_3^2 = 9\sigma_1^2$. What is the P matrix? What does weighted least squares look like?

$$\Omega = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 9 \end{bmatrix}$$

$$P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{3} \end{bmatrix}$$

    The first observations are multiplied by 1, the second observations are multiplied by $\frac{1}{2}$, and the third observations are multiplied by $\frac{1}{3}$.

Clustering

12. What happens when each observation has the same variance, but we instead observe only the average observation over groups of different size?

    The averaged observations have differing variance based on their group size.

13. Explain how to construct the GLS estimator using "clustered" data (data that has been averaged over groups of known size.

    The Ω matrix will be:

$$\Omega = \begin{bmatrix} \frac{1}{n_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{n_m} \end{bmatrix}$$

    where m are the number of groups and n_j is the group size. The P matrix is then:

$$P = \begin{bmatrix} \sqrt{n_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sqrt{n_m} \end{bmatrix}$$

    Each observation $\bar{y}_i$ and $\bar{X}_i$ are multiplied by the square root of their group size.

FGLS

14. Explain what FGLS is.

    FGLS is when we need to estimate the Ω matrix.