LECTURE 9
HETEROSKEDASTICITY AND GENERALIZED LS

Generalized LS

In this lecture, we consider the same model as in Lecture 8, defined by Assumptions (A1), (A6) - (A9). However, we will assume that

(A2)** $E(U_i|X_i) = 0$.

The assumption is stronger than the one we needed for consistency and asymptotic normality of the OLS. The stronger assumption allows us to investigate the issue of efficiency in the case of heteroskedastic errors: $E\left(U_i^2|X_i\right) = \sigma_i^2$, where σ_i^2 is a function of X_i: $\sigma_i^2 = \sigma^2(X_i)$.

Example: suppose that $Y_{i,j} = X_{i,j}'\beta + U_{i,j}$ for $i = 1,\ldots,n$ (n industries) and $j = 1,\ldots,m_i$ (m_i firms in the i-th industry). Assume that the observations are iid across i's and j's. Suppose that the econometrician observes only the average values for the n industries: $\overline{Y}_i = \sum_{j=1}^{m_i} Y_{i,j}/m_i$ and $\overline{X}_i = \sum_{j=1}^{m_i} X_{i,j}/m_i$. Assume that the errors $U_{i,j}$ are homoskedastic, i.e. $E\left(U_{i,j}^2|X_{i,j}\right) = \sigma^2$ for all $i = 1,\ldots,n$ and $j = 1,\ldots,m_i$. However, $\overline{U}_i = \sum_{j=1}^{m_i} U_{i,j}/m_i$, and $E\left(\overline{U}_i^2|\overline{X}_i\right) = \sigma^2/m_i$.

Under heteroskedasticity, the OLS estimator is consistent and asymptotically normal, however, not efficient. There exists an estimator with a smaller (asymptotic) variance. Under (A2**) and (A6), $E\left(\widehat{\beta}_n|X\right) = \beta$ (unbiased), and

$$\begin{aligned} Var\left(\widehat{\beta}_n|X\right) &= \left(X'X\right)^{-1} X'DX \left(X'X\right)^{-1}, \text{ where} \\ D &= \begin{pmatrix} \sigma_1^2 & 0 & \ldots & 0 \\ 0 & \sigma_2^2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & \ldots & 0 & \sigma_n^2 \end{pmatrix}. \end{aligned}$$

Suppose further that $\sigma_i^2 = \sigma^2(X_i)$ is known for all i's. The Generalized LS estimator (GLS) is defined as

$$\begin{aligned} \widehat{\beta}_n^{GLS} &= \left(X'D^{-1}X\right)^{-1} X'D^{-1}Y \\ &= \left(\sum_{i=1}^n \sigma_i^{-2} X_i X_i'\right)^{-1} \sum_{i=1}^n \sigma_i^{-2} X_i Y_i. \end{aligned} \tag{1}$$

In the case of diagonal matrix D, the GLS estimator is also called the *Weighted* LS estimator, since it involves the weighted averages of X_iX_i' and X_iY_i with the weights equal to σ_i^{-2}. Under the assumption $E(U_i|X_i) = 0$ and (A6), we have that $\widehat{\beta}_n^{GLS}$ is unbiased:

$$\begin{aligned} E\left(\widehat{\beta}_n^{GLS}|X\right) &= \beta + \left(X'D^{-1}X\right)^{-1} X'D^{-1}E(U|X) \\ &= \beta. \end{aligned}$$

Its variance is given by

$$\begin{aligned} Var\left(\widehat{\beta}_n^{GLS}|X\right) &= \left(X'D^{-1}X\right)^{-1} X'D^{-1}E\left(UU'|X\right) D^{-1}X \left(X'D^{-1}X\right)^{-1} \\ &= \left(X'D^{-1}X\right)^{-1}. \end{aligned}$$

We will show next that $Var\left(\widehat{\beta}_n^{GLS}|X\right) \le Var\left(\widehat{\beta}_n|X\right)$. First,

$$Var\left(\widehat{\beta}_n^{GLS}|X\right) \le Var\left(\widehat{\beta}_n|X\right) \Leftrightarrow \left(Var\left(\widehat{\beta}_n^{GLS}|X\right)\right)^{-1} \ge \left(Var\left(\widehat{\beta}_n|X\right)\right)^{-1}.$$

Next,

$$\begin{aligned} & X'D^{-1}X - X'X\left(X'DX\right)^{-1}X'X \\ = \quad & X'D^{-1/2}\left(I - D^{1/2}X\left(X'DX\right)^{-1}X'D^{1/2}\right)D^{-1/2}X. \end{aligned}$$

Note that $I - D^{1/2}X\left(X'DX\right)^{-1}X'D^{1/2}$ is a symmetric positive definite matrix, and consequently,

$$X'D^{-1}X \geq X'X\left(X'DX\right)^{-1}X'X.$$

The efficiency of the GLS is actually implied by the Gauss-Markov Theorem. Heteroskedasticity violates one of the assumptions of the Gauss-Markov Theorem. However, consider the transformed model:

$$\begin{aligned} Y_i/\sigma_i &= \left(X_i/\sigma_i\right)'\beta + U_i/\sigma_i \\ Y_i^* &= \left(X_i^*\right)'\beta + U_i^*, \end{aligned}$$

where $Y_i^* = Y_i/\sigma_i$, $X_i^* = X_i/\sigma_i$ and $U_i^* = U_i/\sigma_i$. The transformed residuals U_i^*'s are homoskedastic:

$$\begin{aligned} E\left(\left(U_i^*\right)^2|X_i\right) &= \sigma_i^2/\sigma_i^2 \\ &= 1. \end{aligned}$$

Therefore, by the Gauss-Markov Theorem, the BLUE is given by

$$\begin{aligned} & \left(\sum_{i=1}^n X_i^*\left(X_i^*\right)'\right)^{-1}\sum_{i=1}^n X_i^*Y_i^* \\ = \quad & \left(\sum_{i=1}^n \sigma_i^{-2}X_iX_i'\right)^{-1}\sum_{i=1}^n \sigma_i^{-2}X_iY_i \\ = \quad & \widehat{\beta}_n^{GLS}. \end{aligned}$$

Large sample property of the GLS

We discuss consistency first. Write

$$\widehat{\beta}_n^{GLS} = \beta + \left(\sum_{i=1}^n \sigma_i^{-2}X_iX_i'\right)^{-1}\sum_{i=1}^n \sigma_i^{-2}X_iU_i.$$

We will assume that the function $\sigma^2\left(X_i\right)$ is bounded from below, i.e. with probability one, $\sigma^2\left(X_i\right) \geq \underline{\sigma}^2 > 0$. This is to ensure that $E\left(\sigma_i^{-2}X_iX_i'\right)$ does not "explode". For $r, s = 1, \ldots, k$, we have $E\left|\sigma_i^{-2}X_{i,r}X_{i,s}\right| \leq \underline{\sigma}^{-2}E\left|X_{i,r}X_{i,s}\right| < \infty$ (using Assumption (A7)). By the WLLN and Slutsky's Theorem,

$$\left(n^{-1}\sum_{i=1}^n \sigma_i^{-2}X_iX_i'\right)^{-1} \to_p \left(E\sigma_i^{-2}X_iX_i'\right)^{-1}.$$

Next,

$$\begin{aligned} E\left(\sigma_i^{-2}X_iU_i\right) &= E\left(\sigma_i^{-2}X_iE\left(U_i|X_i\right)\right) \\ &= 0, \end{aligned}$$

and

$$n^{-1}\sum_{i=1}^n \sigma_i^{-2}X_iU_i \to_p 0.$$

Therefore, $\widehat{\beta}_n^{GLS} \to_p \beta$ as $n \to \infty$. Note that, in general, $\widehat{\beta}_n^{GLS}$ is not consistent under (A2*). Since σ_i^2 is a function of X_i, we cannot guarantee that $E\left(\sigma_i^{-2} X_i U_i\right) = 0$ given only $E\left(X_i U_i\right) = 0$.

We show asymptotic normality next. Write

$$n^{1/2}\left(\widehat{\beta}_n^{GLS} - \beta\right) = \left(n^{-1}\sum_{i=1}^n \sigma_i^{-2} X_i X_i'\right)^{-1} n^{-1/2}\sum_{i=1}^n \sigma_i^{-2} X_i U_i.$$

We have

$$\begin{aligned} Var\left(\sigma_i^{-2} X_i U_i\right) &= E\left(\sigma_i^{-4} X_i X_i' U_i^2\right) \\ &= E\left(\sigma_i^{-4} X_i X_i' E\left(U_i^2 | X_i\right)\right) \\ &= E\left(\sigma_i^{-2} X_i X_i'\right). \end{aligned}$$

Hence,

$$\begin{aligned} n^{1/2}\left(\widehat{\beta}_n^{GLS} - \beta\right) &\to_d \left(E\sigma_i^{-2} X_i X_i'\right)^{-1} N\left(0, E\sigma_i^{-2} X_i X_i'\right) \\ &= N\left(0, \left(E\sigma_i^{-2} X_i X_i'\right)^{-1}\right). \end{aligned}$$

Feasible GLS

The GLS estimator is infeasible since σ_i^2 is unknown. A natural solution is to replace unknown σ_i^2 in (1) with their estimates, $\widehat{\sigma}_i^2$. Suppose that σ_i^2 takes the following form:

$$\sigma_i^2 = Z_i'\alpha, \tag{2}$$

where Z_i is some $q \times 1$ function of X_i. Usually, it is assumed that Z_i consists of products and cross-products of the elements of X_i, and a vector of constants (ones). Since $\sigma_i^2 = E\left(U_i^2 | X_i\right)$, we can write

$$U_i^2 = Z_i'\alpha + \nu_i,$$

where $E\left(\nu_i | X_i\right) = 0$. The above model is called *skedastic regression.* Since U_i's are unobservable, one has to use fitted residuals from the OLS regression $\widehat{U}_i$'s instead in order to estimate α:

$$\widehat{\alpha}_n = \left(\sum_{i=1}^n Z_i Z_i'\right)^{-1} \sum_{i=1}^n Z_i \widehat{U}_i^2.$$

One can show that $\widehat{\alpha}_n \to_p \alpha$, and $n^{1/2}\left(\widehat{\alpha}_n - \alpha\right) \to_d N\left(0, V_\alpha\right)$, where V_α is the same as if U_i^2 were observable. The *Feasible* GLS estimator is defined as

$$\begin{aligned} \widehat{\beta}_n^{FGLS} &= \left(X'\widehat{D}_n^{-1}X\right)^{-1} X'\widehat{D}_n^{-1}Y \\ &= \left(\sum_{i=1}^n \widehat{\sigma}_i^{-2} X_i X_i'\right)^{-1} \sum_{i=1}^n \widehat{\sigma}_i^{-2} X_i Y_i, \end{aligned}$$

where

$$\hat{D} = \begin{pmatrix} \widehat{\sigma}_1^2 & 0 & \dots & 0 \\ 0 & \widehat{\sigma}_2^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & \widehat{\sigma}_n^2 \end{pmatrix},$$

and

$$\widehat{\sigma}_i^2 = Z_i'\widehat{\alpha}_n.$$

Further, one can show that $\widehat{\beta}_n^{FGLS} \to_p \beta$, and

$$n^{1/2}\left(\widehat{\beta}_n^{FGLS} - \beta\right) \to_d N\left(0, \left(E\sigma_i^{-2}X_iX_i'\right)^{-1}\right), \quad (3)$$

the same as GLS, provided that (2) is correctly specified. The following are the steps for constructing a FGLS estimator:

1. Obtain $\widehat{\beta}_n$, the OLS estimator of β.
2. Construct $\widehat{U}_i = Y_i - X_i'\widehat{\beta}_n$.
3. Regress $\widehat{U}_i^2$ on Z_i to obtain $\widehat{\alpha}_n$.
4. Construct $\widehat{\sigma}_i^2 = Z_i'\widehat{\alpha}_n$.
5. Compute $\widehat{\beta}_n^{FGLS}$.

One of the problems with the above approach is that $\widehat{\sigma}_i^2 = Z_i'\widehat{\alpha}_n$ can be very close to zero or even negative. There is a number of possible solution. First is truncation. Choose $\underline{\sigma}^2 > 0$ and set $\widehat{\sigma}_i^2 = \max\left\{Z_i'\widehat{\alpha}_n, \underline{\sigma}^2\right\}$. Alternatively, one can consider a nonlinear skedastic regression

$$\sigma_i^2 = \exp\left(Z_i'\alpha\right).$$

Then, in the 3-rd step one should regress $\log \widehat{U}_i^2$ on Z_i, and, in step 4, generate $\widehat{\sigma}_i^2 = \exp\left(Z_i'\widehat{\alpha}_n\right)$.

The FGLS procedure relies on two very strong assumptions. First is that the skedastic regression is correctly specified. If it is misspecified, $\widehat{\sigma}_i^2$ provides only an approximation to σ_i^2. In this case, the asymptotic variance in (3) will be of a sandwich form:

$$\left(E\left(\left(Z_i'\alpha\right)^{-1}X_iX_i'\right)\right)^{-1} E\left(\left(Z_i'\alpha\right)^{-2}\sigma_i^2X_iX_i'\right)\left(E\left(\left(Z_i'\alpha\right)^{-1}X_iX_i'\right)\right)^{-1},$$

and the FGLS will perform worse than OLS. Furthermore, if the assumption $E\left(U_i|X_i\right) = 0$ is violated, the GLS and FGLS estimators are inconsistent. While the OLS estimator is less efficient under certain conditions than FGLS, it provides more *robust* estimates.

Testing for heteroskedasticity

In this section, we discuss a test for $H_0 : \sigma_i^2 = \sigma^2$ with probability one for all i's against the heteroskedastic alternative. If the errors are heteroskedastic, the variance of the OLS estimator is $Var\left(\widehat{\beta}_n|X\right) = \left(\sum_{i=1}^n X_iX_i'\right)^{-1}\sum_{i=1}^n \sigma_i^2X_iX_i'\left(\sum_{i=1}^n X_iX_i'\right)^{-1}$. Under H_0 we have that, $\sum_{i=1}^n \sigma_i^2X_iX_i' = \sigma^2\sum_{i=1}^n X_iX_i'$. The matrix X_iX_i' is $k \times k$ and symmetric. The number of unique elements off the main diagonal is given by $\left(k^2-k\right)/2$, and the total number of unique elements is, therefore, $k\left(k+1\right)/2$. Hence, the null hypothesis imposes $k\left(k+1\right)/2$ restrictions. White (1980) shows that, one can test the null by following the steps below:

1. Obtain the OLS estimator of β, $\widehat{\beta}_n$.
2. Construct fitted OLS residuals as $\widehat{U}_i = Y_i - X_i'\widehat{\beta}_n$.
3. Run the *artificial* skedastic regression $\widehat{U}_i^2$ against all products $(X_{1i}^2, \ldots, X_{ki}^2)$ and the cross-products $(X_{1i}X_{2i}, \ldots, X_{1i}X_{ki}, \ldots, X_{k-1,i}X_{ki})$ of the regressors. (Note that the number of regressors is $k\left(k+1\right)/2$). For example, if the model contains an intercept, say, $X_{1i} = 1$, then the artificial skedastic regression is given by

   $$\widehat{U}_i^2 = \alpha_1 + \alpha_2X_{2i} + \ldots + \alpha_kX_{ki} + \alpha_{k+1}X_{2i}^2 + \ldots \alpha_{2k-1}X_{ki}^2 + \alpha_{2k}X_{2i}X_{3i} + \ldots + \alpha_{k(k+1)/2}X_{k-1,i}X_{ki} + \nu_i.$$

4. Obtain R^2 from the skedastic regression in step 3.
5. Reject the null of homoskedastic errors if $nR^2 > \chi^2_{k(k+1)/2}$. Note that the number of degrees of freedom is given by the number of regressors in the skedastic regression *including* the constant.