**LECTURE 9**
**HETEROSKEDASTICITY AND GENERALIZED LS**

# Generalized LS

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

**(A2**)** $\mathrm{E}[U_i \mid X_i] = 0$.

Assumption (A2**) is stronger than the one we needed for consistency and asymptotic normality of OLS. The stronger assumption allows us to investigate the issue of efficiency in the case of heteroskedastic errors: $\mathrm{E}\left[U_i^2 \mid X_i\right] = \sigma_i^2$, where $\sigma_i^2$ is a function of $X_i$: $\sigma_i^2 = \sigma^2(X_i)$.

**Example.** Suppose that $Y_{i,j} = X_{i,j}^\top \beta + U_{i,j}$ for $i = 1, \dots, n$ ($n$ industries) and $j = 1, \dots, m_i$ ($m_i$ firms in the $i$-th industry). Assume that the observations are iid across $i$ and $j$. 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., $\mathrm{E}\left[U_{i,j}^2 \mid X_{i,j}\right] = \sigma^2$ for all $i = 1, \dots, n$ and $j = 1, \dots, m_i$. However, $\overline{U}_i = \sum_{j=1}^{m_i} U_{i,j}/m_i$, and $\mathrm{E}\left[\overline{U}_i^2 \mid \overline{X}_i\right] = \sigma^2/m_i$.

Under heteroskedasticity, the OLS estimator is consistent and asymptotically normal; however, it is not efficient. An estimator with a smaller asymptotic variance exists. Under (A2**) and (A6), $\mathrm{E}\left[\widehat{\beta}_n \mid X\right] = \beta$ (unbiased), and

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

Suppose further that $\sigma_i^2 = \sigma^2(X_i)$ is known for all $i$. The generalized least-squares (GLS) estimator is defined as

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

When $D$ is diagonal, the GLS estimator is also called the *weighted* least-squares (WLS) estimator, since it involves the weighted averages of $X_i X_i^\top$ and $X_i Y_i$ with the weights equal to $\sigma_i^{-2}$. Under assumption (A2**) and (A6), $\widehat{\beta}_n^{GLS}$ is unbiased:

$$\begin{aligned}
\mathrm{E}\left[\widehat{\beta}_n^{GLS} \mid X\right] &= \beta + \left(X^\top D^{-1} X\right)^{-1} X^\top D^{-1} \mathrm{E}[U \mid X] \\
&= \beta.
\end{aligned}$$

Its variance is given by

$$\begin{aligned}
\mathrm{Var}(\widehat{\beta}_n^{GLS} \mid X) &= \left(X^\top D^{-1} X\right)^{-1} X^\top D^{-1} \mathrm{E}\left[U U^\top \mid X\right] D^{-1} X \left(X^\top D^{-1} X\right)^{-1} \\
&= \left(X^\top D^{-1} X\right)^{-1}.
\end{aligned}$$

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

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

Next,

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

The matrix $I - D^{1/2} X \left(X^\top D X\right)^{-1} X^\top D^{1/2}$ is symmetric and positive semidefinite, and consequently,

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

The efficiency of the GLS estimator is implied by the Gauss–Markov theorem. Heteroskedasticity violates one of the assumptions of the Gauss–Markov theorem. However, consider the transformed model:

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

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

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

By the Gauss–Markov theorem, the BLUE is

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

# Large-sample properties of the GLS

We discuss consistency first. Write

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

Assume that the function $\sigma^2(X_i)$ is bounded from below, i.e., $\sigma^2(X_i) \geq \underline{\sigma}^2 > 0$ almost surely. This is to ensure that $\mathrm{E}\left[\sigma_i^{-2} X_i X_i^\top\right]$ remains finite. For $r, s = 1, \ldots, k$, we have $\mathrm{E}\left|\sigma_i^{-2} X_{i,r} X_{i,s}\right| \leq \underline{\sigma}^{-2}\, \mathrm{E}\left|X_{i,r} X_{i,s}\right| < \infty$ by assumption (A7). By the WLLN and Slutsky's theorem,

$$
\left(n^{-1} \sum_{i=1}^n \sigma_i^{-2} X_i X_i^\top\right)^{-1} \to_p \left(\mathrm{E}[\sigma_i^{-2} X_i X_i^\top]\right)^{-1}.
$$

Next,

$$
\begin{aligned}
\mathrm{E}\left[\sigma_i^{-2} X_i U_i\right] &= \mathrm{E}\left[\sigma_i^{-2} X_i\, \mathrm{E}\left[U_i \mid X_i\right]\right] \\
&= 0,
\end{aligned}
$$

and

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

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

We show asymptotic normality next. Write

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

We have

$$\begin{aligned} \mathrm{Var}(\sigma_i^{-2} X_i U_i) &= \mathrm{E}\left[\sigma_i^{-4} X_i X_i^\top U_i^2\right] \\ &= \mathrm{E}\left[\sigma_i^{-4} X_i X_i^\top \mathrm{E}\left[U_i^2 \mid X_i\right]\right] \\ &= \mathrm{E}\left[\sigma_i^{-2} X_i X_i^\top\right]. \end{aligned}$$

Hence,

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

## Feasible GLS

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

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

where $Z_i$ is a $q \times 1$ vector-valued function of $X_i$. Typically, $Z_i$ consists of products and cross-products of the elements of $X_i$ and a vector of ones. Since $\sigma_i^2 = \mathrm{E}\left[U_i^2 \mid X_i\right]$, we can write

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

where $\mathrm{E}\left[\nu_i \mid X_i\right] = 0$. The above model is called the *skedastic regression*. Since the $U_i$'s are unobservable, one has to use the OLS fitted residuals $\widehat{U}_i$ instead to estimate $\alpha$:

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

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

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

where

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

and

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

Further, $\widehat{\beta}_n^{FGLS} \to_p \beta$, and

$$n^{1/2}\left(\widehat{\beta}_n^{FGLS} - \beta\right) \to_d N\left(0, \left(\mathrm{E}[\sigma_i^{-2} X_i X_i^\top]\right)^{-1}\right), \tag{3}$$

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 $\beta$.
2. Construct $\widehat{U}_i = Y_i - X_i^\top \widehat{\beta}_n$.
3. Regress $\widehat{U}_i^2$ on $Z_i$ to obtain $\widehat{\alpha}_n$.
4. Construct $\widehat{\sigma}_i^2 = Z_i^\top \widehat{\alpha}_n$.
5. Compute $\widehat{\beta}_n^{FGLS}$.

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

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

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

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

$$\left(\mathrm{E}\left[\left(Z_i^\top \alpha\right)^{-1} X_i X_i^\top\right]\right)^{-1} \mathrm{E}\left[\left(Z_i^\top \alpha\right)^{-2} \sigma_i^2 X_i X_i^\top\right] \left(\mathrm{E}\left[\left(Z_i^\top \alpha\right)^{-1} X_i X_i^\top\right]\right)^{-1},$$

and the FGLS may perform worse than the OLS. Furthermore, if the assumption $\mathrm{E}\left[U_i \mid X_i\right] = 0$ is violated, the GLS and FGLS estimators are inconsistent. Although the OLS estimator is less efficient than the FGLS estimator under correct specification, 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 $i$ against the heteroskedastic alternative. If the errors are heteroskedastic, the variance of the OLS estimator is $\mathrm{Var}(\widehat{\beta}_n \mid X) = \left(\sum_{i=1}^n X_i X_i^\top\right)^{-1} \sum_{i=1}^n \sigma_i^2 X_i X_i^\top \left(\sum_{i=1}^n X_i X_i^\top\right)^{-1}$. Under $H_0$, we have that $\sum_{i=1}^n \sigma_i^2 X_i X_i^\top = \sigma^2 \sum_{i=1}^n X_i X_i^\top$. The matrix $X_i X_i^\top$ 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 $\beta$, $\widehat{\beta}_n$.
2. Construct fitted OLS residuals as $\widehat{U}_i = Y_i - X_i^\top \widehat{\beta}_n$.
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. (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

   $$\widehat{U}_i^2 = \alpha_1 + \alpha_2 X_{2i} + \ldots + \alpha_k X_{ki} + \alpha_{k+1} X_{2i}^2 + \ldots + \alpha_{2k-1} X_{ki}^2 + \alpha_{2k} X_{2i} X_{3i} + \ldots + \alpha_{k(k+1)/2} X_{k-1,i} X_{ki} + \nu_i.$$

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