

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

$$\mathbf{(A2^{**})} \quad \mathbb{E}[U_i | X_i] = 0.$$

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: $\mathbb{E}[U_i^2 | X_i] = \sigma_i^2$, where σ_i^2 is a function of X_i : $\sigma_i^2 = \sigma^2(X_i)$.

Example. Suppose that $Y_{i,j} = X_{i,j}^\top \beta + U_{i,j}$ 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: $\bar{Y}_i = \sum_{j=1}^{m_i} Y_{i,j}/m_i$ and $\bar{X}_i = \sum_{j=1}^{m_i} X_{i,j}/m_i$. Assume that the errors $U_{i,j}$ are homoskedastic, that is, $\mathbb{E}[U_{i,j}^2 | X_{i,j}] = \sigma^2$ for all $i = 1, \dots, n$ and $j = 1, \dots, m_i$. However, $\bar{U}_i = \sum_{j=1}^{m_i} U_{i,j}/m_i$, and $\mathbb{E}[\bar{U}_i^2 | \bar{X}_i] = \sigma^2/m_i$.

Under heteroskedasticity, the OLS estimator remains consistent and asymptotically normal. However, it is no longer efficient: the GLS estimator, defined below, achieves a smaller asymptotic variance.

Under (A2^{**}) and (A6), $\mathbb{E}[\hat{\beta}_n | X] = \beta$ (unbiased), and

$$\text{Var}(\hat{\beta}_n | X) = (X^\top X)^{-1} X^\top D X (X^\top X)^{-1}, \text{ where}$$

$$D = \begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n^2 \end{pmatrix}.$$

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

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

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 σ_i^{-2} . Under Assumptions (A2^{**}) and (A6), $\hat{\beta}_n^{GLS}$ is unbiased:

$$\begin{aligned} \mathbb{E}[\hat{\beta}_n^{GLS} | X] &= \beta + (X^\top D^{-1} X)^{-1} X^\top D^{-1} \mathbb{E}[U | X] \\ &= \beta. \end{aligned}$$

Its variance is given by

$$\begin{aligned} \text{Var}(\hat{\beta}_n^{GLS} | X) &= (X^\top D^{-1} X)^{-1} X^\top D^{-1} \mathbb{E}[U U^\top | X] D^{-1} X (X^\top D^{-1} X)^{-1} \\ &= (X^\top D^{-1} X)^{-1}. \end{aligned}$$

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

$$\text{Var}(\widehat{\beta}_n^{GLS} | X) \leq \text{Var}(\widehat{\beta}_n | X) \Leftrightarrow (\text{Var}(\widehat{\beta}_n^{GLS} | X))^{-1} \geq (\text{Var}(\widehat{\beta}_n | X))^{-1}.$$

Next,

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

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

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

The efficiency of the GLS estimator can also be understood through the Gauss–Markov theorem applied to a transformed model. Heteroskedasticity violates the homoskedasticity assumption of the Gauss–Markov theorem, but transforming the data restores it:

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

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:

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

By the Gauss–Markov theorem, the BLUE is

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

Large-sample properties of the GLS

We establish consistency first. Substituting the true model into the GLS formula gives

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

Assume that $\sigma^2(X_i) \geq \underline{\sigma}^2 > 0$ almost surely for some constant $\underline{\sigma}^2$. This lower bound ensures that $\mathbb{E}[\sigma_i^{-2} X_i X_i^\top]$ remains finite. For $r, s = 1, \dots, k$, we have $\mathbb{E}[\sigma_i^{-2} X_{i,r} X_{i,s}] \leq \underline{\sigma}^{-2} \mathbb{E}[X_{i,r} X_{i,s}] < \infty$ by assumption (A7). By the WLLN and Slutsky's theorem,

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

Next,

$$\begin{aligned} \mathbb{E}[\sigma_i^{-2} X_i U_i] &= \mathbb{E}[\sigma_i^{-2} X_i \mathbb{E}[U_i | X_i]] \\ &= 0, \end{aligned}$$

and

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

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

Turning to asymptotic normality, rescale the estimation error as

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

We have

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

Hence,

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

Feasible GLS

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

$$\sigma_i^2 = Z_i^\top \alpha, \quad (2)$$

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 $\sigma_i^2 = E[U_i^2 | X_i]$, we can write

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

where $E[\nu_i | X_i] = 0$. This equation is called the *skedastic regression*. Since the errors U_i are unobservable, one replaces U_i^2 with the squared OLS residuals \widehat{U}_i^2 to estimate α :

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

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

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

where

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

and

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

Further, $\hat{\beta}_n^{FGLS} \rightarrow_p \beta$, and

$$n^{1/2} \left(\hat{\beta}_n^{FGLS} - \beta \right) \rightarrow_d N \left(0, \left(\mathbb{E}[\sigma_i^{-2} X_i X_i^\top] \right)^{-1} \right), \quad (3)$$

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

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

One of the problems with the above approach is that $\hat{\sigma}_i^2 = Z_i^\top \hat{\alpha}_n$ 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 $\hat{\sigma}_i^2 = \max \{ Z_i^\top \hat{\alpha}_n, \underline{\sigma}^2 \}$. 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 $\log \hat{U}_i^2$ on Z_i , and, in step 4, generate $\hat{\sigma}_i^2 = \exp \left(Z_i^\top \hat{\alpha}_n \right)$.

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

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

and the FGLS may perform worse than OLS.

Second, the conditional mean assumption $\mathbb{E}[U_i | X_i] = 0$ must hold. If it is violated, both GLS and FGLS are inconsistent, whereas OLS remains consistent under the weaker condition $\mathbb{E}[X_i U_i] = 0$. Thus, although OLS is less efficient than FGLS under correct specification, it provides more *robust* estimates.

Testing for heteroskedasticity

We now present a test of the null hypothesis $H_0 : \sigma_i^2 = \sigma^2$ for all i (homoskedasticity) against the alternative that error variances differ across observations. If the errors are heteroskedastic, the variance of the OLS estimator is $\text{Var}(\hat{\beta}_n | X) = \left(\sum_{i=1}^n X_i X_i^\top \right)^{-1} \sum_{i=1}^n \sigma_i^2 X_i X_i^\top \left(\sum_{i=1}^n X_i X_i^\top \right)^{-1}$. 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, so it has $k(k+1)/2$ unique elements. White (1980) shows that one can test the null by following the steps below:

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

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

4. Obtain R^2 from the skedastic regression in step 3.
5. Reject the null of homoskedasticity if $nR^2 > \chi_{k(k+1)/2-1, 1-\alpha}^2$. Under H_0 , all slope coefficients in the skedastic regression are zero. A general result states that if the true slopes in a linear regression with q regressors and a constant are all zero, then $nR^2 \rightarrow_d \chi_q^2$. The degrees of freedom equal the number of regressors in the skedastic regression *excluding* the constant, which is $k(k+1)/2 - 1$.