Chapter 7

Generalized and Weighted Least Squares Estimation

The usual linear regression model assumes that all the random error components are identically and independently distributed with constant variance. When this assumption is violated, then ordinary least squares estimator of the regression coefficient loses its property of minimum variance in the class of linear and unbiased estimators. The violation of such assumption can arise in anyone of the following situations:

1. The variance of random error components is not constant.
2. The random error components are not independent.
3. The random error components do not have constant variance as well as they are not independent.

In such cases, the covariance matrix of random error components does not remain in the form of an identity matrix but can be considered as any positive definite matrix. Under such assumption, the OLSE does not remain efficient as in the case of an identity covariance matrix. The generalized or weighted least squares method is used in such situations to estimate the parameters of the model.

In this method, the deviation between the observed and expected values of y_i is multiplied by a weight ω_i where ω_i is chosen to be inversely proportional to the variance of y_i.

For a simple linear regression model, the weighted least squares function is

$$S(\beta_0, \beta_1) = \sum^{n} \omega_i \left(y_i - \beta_0 - \beta_1 x_i \right)^2 .$$

The least-squares normal equations are obtained by differentiating $S(\beta_0, \beta_1)$ with respect to β_0 and β_1 and equating them to zero as

$$\hat{\beta}_0 \sum_{i=1}^{n} \omega_i + \hat{\beta}_1 \sum_{i=1}^{n} \omega_i x_i = \sum_{i=1}^{n} \omega_i y_i$$

$$\hat{\beta}_0 \sum_{i=1}^{n} \omega_i x_i + \hat{\beta}_1 \sum_{i=1}^{n} \omega_i x_i^2 = \sum_{i=1}^{n} \omega_i x_i y_i .$$

The solution of these two normal equations gives the weighted least squares estimate of β_0 and β_1.

Generalized least squares estimation

Suppose in usual multiple regression model

$$y = X\beta + \varepsilon \text{ with } E(\varepsilon) = 0, V(\varepsilon) = \sigma^2 I,$$

the assumption $V(\varepsilon) = \sigma^2 I$ is violated and become

$$V(\varepsilon) = \sigma^2 \Omega$$

where Ω is a known $n \times n$ nonsingular, positive definite and symmetric matrix.

This structure of Ω incorporates both the cases.

- when Ω is diagonal but with unequal variances and
- when Ω is not necessarily diagonal depending on the presence of correlated errors, some of the diagonal elements are nonzero.

The OLSE of β is

$$b = (X'X)^{-1} X'y$$

In such cases, OLSE gives unbiased estimate but has more variability as

$$E(b) = (X'X)^{-1} X'E(y) = (X'X)^{-1} X'X\beta = \beta$$
$$V(b) = (X'X)^{-1} X'V(y)X(X'X)^{-1} = \sigma^2 (X'X)^{-1} X'\Omega X(X'X)^{-1}.$$

Now we attempt to find better estimator as follows:

Since Ω is positive definite, symmetric, so there exists a nonsingular matrix K such that.

$$KK' = \Omega.$$

Then in the model

$$y = X\beta + \varepsilon,$$

premutliply by K^{-1}, this gives

$$K^{-1}y = K^{-1}X\beta + K^{-1}\varepsilon$$

or

$$z = B\beta + g$$

where $z = K^{-1}y, B = K^{-1}X, g = K^{-1}\varepsilon$. Now observe that

$$E(g) = K^{-1)}E(\varepsilon) = 0$$

and

$$
\begin{aligned}
V(g) &= E\left[\{g - E(g)\}\{g - E(g)\}'\right] \\
&= E(gg') \\
&= E\left[K^{-1}\varepsilon\varepsilon' K'^{-1}\right] \\
&= K^{-1}E(\varepsilon\varepsilon')K'^{-1} \\
&= \sigma^2 K^{-1}\Omega K'^{-1} \\
&= \sigma^2 K^{-1}KK'K'^{-1} \\
&= \sigma^2 I.
\end{aligned}
$$

Thus the elements of g have 0 mean, and they are uncorrelated.

So either minimize $S(\beta) = g'g$

$$
\begin{aligned}
&= \varepsilon'\Omega^{-1}\varepsilon \\
&= (y - X\beta)'\Omega^{-1}(y - X\beta)
\end{aligned}
$$

and get normal equations as

$$(X'\Omega^{-1}X)\hat{\beta} = X'\Omega^{-1}y$$

or $\hat{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y.$

Alternatively, we can apply OLS to transformed model and obtain OLSE of β as

$$
\begin{aligned}
\hat{\beta} &= (B'B)^{-1}B'z \\
&= (X'K'^{-1}K^{-1}X)^{-1}X'K'^{-1}K^{-1}y \\
&= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y
\end{aligned}
$$

This is **termed as generalized least squares estimator (GLSE**) of β.

The estimation error of GLSE is

$$
\begin{aligned}
\hat{\beta} &= (B'B)^{-1}B'(B\beta + g) \\
&= \beta + (B'B)^{-1}B'g
\end{aligned}
$$

or $\hat{\beta} - \beta = (B'B)^{-1}B'g.$

Then

$$E(\hat{\beta} - \beta) = (B'B)^{-1}B'E(g) = 0$$

which shows that GLSE is an unbiased estimator of β. The covariance matrix of GLSE is given by

$$
\begin{aligned}
V(\hat{\beta}) &= E\left[\{\hat{\beta} - E(\hat{\beta})\}\{\hat{\beta} - E(\hat{\beta})\}'\right] \\
&= E\left[(B'B)^{-1}B'gg'B'(B'B)^{-1}\right] \\
&= (B'B)^{-1}B'E(gg')B'(B'B)^{-1}.
\end{aligned}
$$

Since

$$
\begin{aligned}
E(gg') &= K^{-1}E(\varepsilon\varepsilon')K'^{-1} \\
&= \sigma^2 K^{-1}\Omega K'^{-1} \\
&= \sigma^2 K^{-1}KK'K'^{-1} \\
&= \sigma^2 I,
\end{aligned}
$$

so

$$
\begin{aligned}
V(\hat{\beta}) &= \sigma^2 (B'B)^{-1}B'B(B'B)^{-1} \\
&= \sigma^2 (B'B)^{-1} \\
&= \sigma^2 (X'K'^{-1}K^{-1}X)^{-1} \\
&= \sigma^2 (X'\Omega^{-1}X)^{-1}.
\end{aligned}
$$

Now we prove that GLSE is the best linear unbiased estimator of β.

The Gauss-Markov theorem for the case $Var(\varepsilon) = \Omega$

The Gauss-Markov theorem establishes that the generalized least-squares (GLS) estimator of β given by $\hat{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y,$ is BLUE (best linear unbiased estimator). By best β, we mean that $\hat{\beta}$ minimizes the variance for any linear combination of the estimated coefficients, $\ell'\hat{\beta}$. We note that

$$
\begin{aligned}
E(\hat{\beta}) &= E\left[(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y\right] \\
&= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}E(y) \\
&= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}X\beta \\
&= \beta.
\end{aligned}
$$

Thus $\hat{\beta}$ is an unbiased estimator of β.

The covariance matrix of $\hat{\beta}$ is given by

$$
\begin{aligned}
V(\hat{\beta}) &= \left[(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\right]V(y)\left[(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\right]' \\
&= \left[(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\right]\Omega\left[(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\right]' \\
&= \left[(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\right]\Omega\left[\Omega^{-1}X(X'\Omega^{-1}X)^{-1}\right] \\
&= (X'\Omega^{-1}X)^{-1}.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
Var(\ell'\hat{\beta}) &= \ell' Var(\hat{\beta})\ell \\
&= \ell'\left[(X'\Omega^{-1}X)^{-1}\right]\ell.
\end{aligned}
$$

Let $\tilde{\beta}$ be another unbiased estimator of β that is a linear combination of the data. Our goal, then, is to show that $Var(\ell'\tilde{\beta}) \geq \ell'(X'\Omega^{-1}X)^{-1}\ell$ with at least one ℓ such that $Var(\ell'\tilde{\beta}) \geq \ell'(X'\Omega^{-1}X)^{-1}\ell$.

We first note that we can write any other estimator of β that is a linear combination of the data as

$$\tilde{\beta} = \left[(X'\Omega^{-1}X)^{-1}X'\Omega^{-1} + B\right]y + b_0^*$$

where B is an $p \times n$ matrix and b_o^* is a $p \times 1$ vector of constants that appropriately adjusts the GLS estimator to form the alternative estimate. Then

$$\begin{aligned}
E(\tilde{\beta}) &= E\left(\left[(X'\Omega^{-1}X)^{-1}X'\Omega^{-1} + B\right]y + b_o^*\right) \\
&= \left[(X'\Omega^{-1}X)^{-1}X'\Omega^{-1} + B\right]E(y) + b_0^* \\
&= \left[(X'\Omega^{-1}X)^{-1}X'\Omega^{-1} + B\right]XB + b_0^* \\
&= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}X\beta + BX\beta + b_0^* \\
&= \beta + BX\beta + b_0^*.
\end{aligned}$$

Consequently, $\tilde{\beta}$ is unbiased if and only if both $b_0^* = 0$ and $BX = 0$. The covariance matrix of $\tilde{\beta}$ is

$$\begin{aligned}
V(\tilde{\beta}) &= Var\left(\left[(X'\Omega^{-1}X)^{-1}X'\Omega^{-1} + B\right]y\right) \\
&= \left[(X'\Omega^{-1}X)^{-1}X'\Omega^{-1} + B\right]V(y)\left[(X'\Omega^{-1}X)^{-1}X'\Omega^{-1} + B\right]' \\
&= \left[(X'\Omega^{-1}X)^{-1}X'\Omega^{-1} + B\right]\Omega\left[(X'\Omega^{-1}X)^{-1}X'\Omega^{-1} + B\right]' \\
&= \left[(X'\Omega^{-1}X)^{-1}X'\Omega^{-1} + B\right]\Omega\left[\Omega^{-1}X(X'\Omega^{-1}X)^{-1} + B'\right] \\
&= \left[(X'\Omega^{-1}X)^{-1} + B\Omega B'\right]
\end{aligned}$$

because $BX = 0$, which implies that $(BX)' = X'B' = 0$. Then

$$\begin{aligned}
Var(\ell'\tilde{\beta}) &= \ell'V(\tilde{\beta})\ell \\
&= \ell'\left(\left[(X'\Omega^{-1}X)^{-1} + B\Omega B'\right]\right)\ell \\
&= \ell'(X'\Omega^{-1}X)^{-1}\ell + \ell'B\Omega B'\ell \\
&= Var(\ell'\hat{\beta}) + \ell'B\Omega B'\ell.
\end{aligned}$$

We note that Ω is a positive definite matrix. Consequently, there exists some nonsingular matrix K such that $\Omega = K'K$ As a result, $B\Omega B' = BK'KB'$ is at least positive semidefinite matrix; hence, $\ell'B\Omega B'\ell \geq 0$. Next note that we can define $\ell^* = KB'\ell$. As a result,

$$\ell'B\Omega B'\ell = \ell^{*'}\ell^* = \sum_{i=1}^{p} \ell_i^{*2}$$

which must be strictly greater than 0 for some $\ell \neq 0$ unless $B = 0$. Thus, the GLS estimate of β is the best linear unbiased estimator.

Weighted least squares estimation

When $\varepsilon's$ are uncorrelated and have unequal variances, then

$$V(\varepsilon)=\sigma^2\Omega=\sigma^2\begin{bmatrix} \frac{1}{\omega_1} & 0 & 0 & \cdots & 0 \\ 0 & \frac{1}{\omega_2} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{1}{\omega_n} \end{bmatrix}.$$

The estimation procedure is usually called as weighted least squares.

Let $W=\Omega^{-1}$ then the weighted least squares estimator of β is obtained by solving normal equation $(X'WX)\hat{\beta}=X'Wy$ which gives

$$\hat{\beta}=(X'WX)^{-1}X'Wy$$

where $\omega_1,\omega_2,...,\omega_n$ are called the **weights**.

The observations with large variances usual have smaller weights than observations with small variance.