# LECTURE 11: GENERALIZED LEAST SQUARES (GLS)

In this lecture, we will consider the model $y = X\beta + \varepsilon$ retaining the assumption $Ey = X\beta$.

However, we no longer have the assumption $V(y) = V(\varepsilon) = \sigma^2 I$. Instead we add the assumption $V(y) = V$ where V is positive definite. Sometimes we take $V = \sigma^2\Omega$ with $\text{tr}\,\Omega = N$

As we know, $\hat{\beta} = (X'X)^{-1}X'y$. What is $E\hat{\beta}$?

Note that $V(\hat{\beta}) = (X'X)^{-1}X\,VX(X'X)^{-1}$ in this case.

Is $\hat{\beta}$ BLUE? Does $\hat{\beta}$ minimize $e'e$?

The basic idea behind GLS is to transform the observation matrix [y X] so that the variance in the transformed model is I (or σ2I).

Since V is positive definite, $V^{-1}$ is positive definite too. Therefore, there exists a nonsingular matrix P such that $V^{-1} = P'P$.

Transforming the model $y = X\beta + \varepsilon$ by P yields $Py = PX\beta + P\varepsilon$.

Note that $EP\varepsilon = PE\varepsilon = 0$ and
$V(P\varepsilon) = PE\varepsilon\varepsilon'P' = PVP' = P(P'P)^{-1}P' = I$.
(We could have done this with $V = \sigma^2\Omega$ and imposed tr $\Omega = N$ if useful.)
That is, the transformed model $Py = PX\beta + P\varepsilon$ satisfies the conditions under which we developed our Least Squares estimators.

Thus, the LS estimator is BLUE in the transformed model. The LS estimator for $\beta$ in the model $Py = PX\beta + P\varepsilon$ is referred to as the GLS estimator for $\beta$ in the model $y = X\beta + \varepsilon$.

*Proposition:* The GLS estimator for $\beta$ is $\hat{\beta}_G = (X'V^{-1}X)^{-1}X'V^{-1}y$.

*Proof:* Apply LS to the transformed model. Thus, $\hat{\beta}_G = (X'P'PX)^{-1}X'P'Py = (X'V^{-1}X)^{-1}X'V^{-1}y$ □

*Proposition:* $V(\hat{\beta}_G) = (X'V^{-1}X)^{-1}$

*Proof:* Note that $\hat{\beta}_G - \beta = (X'V^{-1}X)^{-1}X'V^{-1}\varepsilon$. Thus,

$$V(\hat{\beta}_G) = E\,(X'V^{-1}X)^{-1}X'V^{-1}\varepsilon\varepsilon'V^{-1}X(X'V^{-1}X)^{-1}$$
$$= (X'V^{-1}X)^{-1}X'V^{-1}VV^{-1}X(X'V^{-1}X)^{-1} = (X'V^{-1}X)^{-1}$$

*Aitken's Theorem:* The GLS estimator is BLUE. (This really follows from the Gauss-Markov Theorem, but let's give a direct proof.)

*Proof:* Let b be an alternative linear unbiased estimator such that

$b = [(X'V^{-1}X)^{-1}X'V^{-1} + A]y.$
Unbiasedness implies that $AX = 0$.

$$V(b) = [(X'V^{-1}X)^{-1}X'V^{-1} + A]\, V\, [(X'V^{-1}X)^{-1}X'V^{-1} + A]'$$

$$= (X'V^{-1}X)^{-1} + AVA' + (X'V^{-1}X)^{-1}X'A' + AX(X'V^{-1}X)^{-1}$$

The last two terms are zero. (*why?*)

The second term is positive semi-definite, so $A = 0$ is best. □

*What does GLS minimize?*

Recall that $(y - Xb)'(y - Xb)$ is minimized by $b = \hat{\beta}$
(i.e. $(y - Xb)$ is minimized in length by $b = \hat{\beta}$).

Consider $P(y - Xb)$. The length of this vector is $(y - Xb)'P'P(y - Xb) = (y - Xb)'V^{-1}(y - Xb)$.
Thus, GLS minimizes $P(y - Xb)$ in length.

Let $\tilde{e} = (y - X\hat{\beta}_G)$. Note that satisfies $X'V^{-1}(y - X\hat{\beta}_G) = X'V^{-1}\tilde{e} = 0$. (*why?*)

Then

$$(y - Xb)'V^{-1}(y - Xb) = (y - X\hat{\beta}_G)'V^{-1}(y - X\hat{\beta}_G) + (b - \hat{\beta}_G)'X'V^{-1}X(b - \hat{\beta}_G).$$

Note that $X'\tilde{e} \neq 0$ in general.

Estimation of $\sigma^2$:

Let $V(y) = \sigma^2\Omega$ where $\text{tr}\ \Omega = N$.

Choose P so $P'P = \Omega^{-1}$. Then the variance in the transformed model $Py = PX\beta + P\varepsilon$ is $\sigma^2 I$. Our standard formula gives $s^2 = \tilde{e}'\tilde{e}/(N - K)$ which is the unbiased estimator for $\sigma^2$.

Now we add the assumption of normality: $y \sim N(X\beta, \sigma^2\Omega)$.

Consider the log likelihood:

$$\ell(\beta\sigma^2) = c - \frac{N}{2}\ln\sigma^2 - \frac{1}{2}\ln|\Omega| - \frac{1}{2\sigma^2}(y - X\beta)'\Omega^{-1}(y - X\beta)$$

Proposition: The GLS estimator is the ML estimator for $\beta$. (why?)

*Proposition:* $\sigma^2_{ML} = \tilde{e}\tilde{e}'/N$ (*as expected*).

*Proposition:* $\hat{\beta}_G$ and $\tilde{e}$ are independent. (*How would you prove this?*)

*Testing:*

Testing procedures are as in the ordinary model. Results we have developed under the standard set-up will be applied to the transformed model.

*When does* $\hat{\beta}_G = \hat{\beta}$ *?*

1. $\hat{\beta}_G = \hat{\beta}$ holds trivially when $\sigma^2 I = V$.

2. $\hat{\beta} = (X'X)^{-1}X'y$ and $\hat{\beta}_G = (X'V^{-1}X)^{-1}X'V^{-1}y$

$$\hat{\beta}_G = \hat{\beta}$$

$$\Rightarrow (X'X)^{-1}X' = (X'V^{-1}X)^{-1}X'V^{-1}$$

$\Rightarrow VX = X(X'V^{-1}X)^{-1}X'X = XR$ (*What are the dimensions of these matrices?*)

*Interpretation:* In the case where K = 1, X is an eigenvector of V. In general, if the columns of X
are each linear combinations of the same K eigenvectors of V, then $\hat{\beta}_G = \hat{\beta}$. This is hard to check and would usually be a bad assumption.

*Example:* Equicorrelated case:
$V(y) = V = I + \alpha 1 1'$ where 1 is an N-vector of ones.

The LS estimator is the same as the GLS estimator if X has a column of ones

Case of unknown $\Omega$:

Note that there is no hope of estimating $\Omega$ since there are $N(N + 1)/2$ parameters and only N observations. Thus, we usually make some parametric restriction as $\Omega = \Omega(\theta)$ with $\theta$ a fixed parameter. Then we can hope to estimate $\theta$ consistently using squares and cross products of LS residuals or we could use ML.

Note that it doesn't make sense to try to consistently estimate $\Omega$ since it grows with sample size.

Thus, "consistency" refers to the estimate of $\theta$.
*Definition:* $\hat{\Omega} = \Omega(\hat{\theta})$ is a consistent estimator of $\Omega$ if and only if $\hat{\theta}$ is a consistent estimator of $\theta$.

<u>Feasible GLS (FGLS)</u> is the estimation method used when $\Omega$ is unknown. FGLS is the same as GLS except that it uses an estimated $\Omega$, say $\hat{\Omega} = \Omega(\hat{\theta})$, instead of $\Omega$.

*Proposition:* $\hat{\beta}_{FG} = (X' \hat{\Omega}^{-1} X)^{-1} X' \hat{\Omega}^{-1} y$

Note that $\hat{\beta}_{FG} = \beta + (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}\varepsilon$. The following proposition follows easily from this decomposition of $\hat{\beta}_{FG}$

*Proposition:* The <u>sufficient</u> conditions for $\hat{\beta}_{FG}$ to be consistent are

$$plim \frac{X'\hat{\Omega}^{-1}X}{N} = Q$$

where Q is positive definite and finite, and

$$plim \frac{X'\hat{\Omega}^{-1}\varepsilon}{N} = 0.$$

It takes a little more to get a distribution theory. From our discussion of $\hat{\beta}_G$, it easily follows that

$$\sqrt{N}(\hat{\beta}_G - \beta) \rightarrow N\left(0,\ \sigma^2\left(\frac{X'\Omega^{-1}X}{N}\right)^{-1}\right)$$

What about the distribution of $\hat{\beta}_{FG}$ when $\Omega$ is unknown?

*Proposition:* <u>Sufficient</u> conditions for $\hat{\beta}_{FG}$ and $\hat{\beta}_{G}$ to have the same asymptotic distribution are that

$$\mathit{plim}\ \frac{X'(\hat{\Omega}^{-1} - \Omega^{-1})X}{N} = 0$$

$$\mathit{plim}\ \frac{X'(\hat{\Omega}^{-1} - \Omega^{-1})\varepsilon}{\sqrt{N}} = 0.$$

*Proof:* Note that

$$\sqrt{N}(\hat{\beta}_{G} - \beta) = \left(\frac{X'\Omega^{-1}X}{N}\right)^{-1}\left(\frac{X'\Omega^{-1}\varepsilon}{\sqrt{N}}\right)$$

and

$$\sqrt{N}(\hat{\beta}_{FG} - \beta) = \left(\frac{X'\hat{\Omega}^{-1}X}{N}\right)^{-1}\left(\frac{X'\hat{\Omega}^{-1}\varepsilon}{\sqrt{N}}\right).$$

Thus

$$plim \sqrt{N}(\hat{\beta}_G - \hat{\beta}_{FG}) = 0$$

if

$$plim \frac{X'\hat{\Omega}^{-1}X}{N} = plim \frac{X'\Omega^{-1}X}{N}$$

and

$$plim \frac{X'\hat{\Omega}^{-1}\varepsilon}{\sqrt{N}} = plim \frac{X'\Omega^{-1}\varepsilon}{\sqrt{N}}.$$

We are done. (Recall that plim(x−y) = 0 ⇒ the random variables x and y have the same asymptotic distribution.)

*Summing up:*

Consistency of $\hat{\theta}$ implies consistency of the FGLS estimator. A little more is required for the

FGLS estimator to have the same asymptotic distribution as the GLS estimator. These conditions are usually met.

*Small-sample properties of FGLS estimators:*

*Proposition:* Suppose $\hat{\theta}$ is an <u>even</u> function of ε (i.e. (ε) = (-ε)). (This holds if $\hat{\theta}$ depends on squares and cross products of residuals.) Suppose ε has a symmetric distribution. Then $E\hat{\beta}_{FG} = \beta$ if the mean exists.

*Proof:* The sampling error

$$\hat{\beta}_{FG} - \beta = (X'\hat{\Omega}(\hat{\theta})^{-1}X)^{-1}X'\hat{\Omega}(\hat{\theta})^{-1}\varepsilon$$

has a symmetric distribution around zero since ε and -ε give the same value of $\hat{\Omega}$. If the mean exists, it is zero. □

Note that this property is weak . It is easily obtained but it is not very useful.

*General advice:*

-Do not use too many parameters in estimating the variance-covariance matrix or the increase in
sampling variances will outweigh the decrease in asymptotic variance.

-Always calculate LS as well as GLS estimators. What are the data telling you if these differ a lot?