Economics 620, Lecture 11: Generalized Least Squares (GLS)

Nicholas M. Kiefer

Cornell University

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 $tr\Omega = N$.

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

Note that $V(\hat{\beta}) = (X'X)^{-1}XVX(X'X)^{-1}$ 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 $\sigma^2 I$).

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 Least Squares estimators.

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

Proposition: The LGS estimator for β is

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

Proof: Apply LS to the transformed model. Thus,

$$\begin{aligned} \hat{\beta}_G &= (X'P'PX)^{-1}X'P'Py \\ &= (X'V^{-1}X)^{-1}X'V^{-1}y. \end{aligned}$$

■

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,

$$\begin{aligned} V(\hat{\beta}_G) &= E(X'V^{-1}X)^{-1}X'V^{-1}\varepsilon\varepsilon'V^{-1}X(X'V^{-1}X)^{-1} \\ &= (X'V^{-1}X)^{-1}X'V^{-1}VV^{-1}X(X'V^{-1}X)^{-1} \\ &= (X'V^{-1}X)^{-1}. \end{aligned}$$

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 $AX = 0$.

$$\begin{aligned} V(b) &= [(X'V^{-1}X)^{-1}X'V^{-1} + A]V \\ &\quad \times[(X'V^{-1}X)^{-1}X'V^{-1} + A'] \\ &= (X'V^{-1}X)^{-1} + AVA' + (X'V^{-1}X)^{-1}X'A' \\ &\quad +AX(X'V^{-1}X)^{-1} \end{aligned}$$

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

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

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

Estimation of Variance

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

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 σ^2.

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

Consider the log likelihood:

$$\begin{aligned}\ell(\beta\sigma^2) &= c - \frac{N}{2}\ln\sigma^2 - \frac{1}{2}\ln|\Omega| \\ &\quad - \frac{1}{2\sigma^2}(y - X\beta)'\Omega^{-1}(y - X\beta).\end{aligned}$$

Proposition: The GLS estimator is the ML estimator for β. (*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. $$\begin{aligned} \hat{\beta} &= (X'X)^{-1}X'y \text{ and} \\ \hat{\beta}_G &= (X'V^{-1}X)^{-1}X'V^{-1}y \\ \hat{\beta}_G &= \hat{\beta} \\ &\Rightarrow (X'X)^{-1}X' = (X'V^{-1}X)^{-1}X'V^{-1} \\ &\Rightarrow VX = X(X'V^{-1}X)^{-1}X'X = XR \end{aligned}$$

(*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: $V(y) = V = I + \alpha 11'$ 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 Ω:

Note that there is no hope of estimating Ω since there are $N(N+1)/2$ parameters and only N observations. Thus, we usually make some parametric restriction as $\Omega = \Omega(\theta)$ with θ a fixed parameter. Then we can hope to estimate θ 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 Ω since it grows with sample size.

Thus, "consistency" refers to the estimate of θ.

Defintion: $\hat{\Omega} = \Omega(\hat{\theta})$ is a consistent estimator of Ω if and only if $\hat{\theta}$ is a consistent estimator of θ.

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

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 *sufficient* conditions for $\hat{\beta}_{FG}$ to be consistent are

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

where Q is positive definite and finite, and

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

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

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

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

Proposition: **Sufficient** conditions for $\hat{\beta}_{FG}$ and $\hat{\beta}_G$ to have the same asymptotic distribution are that

$$
\begin{aligned}
p\lim \frac{X'(\hat{\Omega}^{-1} - \Omega^{-1})X}{N} &= 0 \\
p\lim \frac{X'(\hat{\Omega}^{-1} - \Omega^{-1})e}{\sqrt{N}} &= 0.
\end{aligned}
$$

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

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

if

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

and

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

We are done. (Recall that $p\lim(x - y) = 0 \Rightarrow$ 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 **even** function of ε (i.e., $\hat{\theta}(\varepsilon) = \hat{\theta}(-\varepsilon)$). (This holds of $\hat{\theta}$ 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 $-\varepsilon$ 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 note 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?