7. GENERALIZED LEAST SQUARES (GLS)

[1] ASSUMPTIONS:

- Assume SIC except that $\text{Cov}(\varepsilon) = \text{E}(\varepsilon\varepsilon') = \sigma^2 \Omega$ where $\Omega \neq I_T$. Assume that $\text{E}(\varepsilon) = 0_{T \times 1}$, and that $X'\Omega^{-1}X$ and $X'\Omega X$ are all positive definite.

Examples:
- Autocorrelation: The $\varepsilon_i$ are serially correlated. ($\Omega$ is not diagonal.)
- Heteroskedasticity: $\Omega$ is diagonal, but diagonals are not identical.

[2] PROPERTIES OF OLS

Theorem: $\hat{\beta}$ is unbiased (and consistent).
Proof: $\hat{\beta} = \beta + (X'X)^{-1}X'\varepsilon \rightarrow \text{E}(\hat{\beta}) = \beta$.

Theorem: $\text{Cov}(\hat{\beta}) = (X'X)^{-1}X'\sigma^2 \Omega X(X'X)^{-1}$.
Proof:

$$\text{Cov}(\hat{\beta}) = \text{E} \left( (\hat{\beta} - \beta)(\hat{\beta} - \beta)' \right) = \text{E} \left( (X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1} \right)$$

$$= (X'X)^{-1}X'\text{E}(\varepsilon\varepsilon')X(X'X)^{-1} = (X'X)^{-1}X'\sigma^2 \Omega X(X'X)^{-1}.$$  

Comment: All the usual $t$ and $F$ tests are invalid. This is because $s^2(X'X)^{-1}$ is no longer an unbiased estimator of $\text{Cov}(\hat{\beta})$. 

GLS-1
**GLS ESTIMATOR**

(3.1) **CASE I: Ω is known**

**Theorem:**
Assume that Ω is positive definite. Then, there exist a T×T nonsingular matrix V, such that V′V = Ω⁻¹.

**Comment:**
For GLS, it is sufficient to find V such that V′V = aΩ⁻¹, where a is some positive constant.

**Theorem:**

\[ VΩV' = I_T \]

**Proof:**

\[ VΩV' = V(V'V)^{-1}V' = VV^{-1}(V')^{-1}V' = I_T \bullet I_T = I_T. \]

**Theorem:**
Assume that \( X'Ω^{-1}X \) is positive definite. Then, \( Vy = VXβ + Vε \) satisfies ideal conditions.

**Proof:**

\[ E(Vε) = VE(ε) = 0_{T×1}. \]
\[ Cov(Vε) = VCov(ε)V' = V\sigma^2ΩV' = \sigma^2I_T. \]
Theorem: (Aitken)

The BLUE of $\beta$ is the GLS estimator $\tilde{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$.

Proof:
Since $V_y = VX\beta + V\varepsilon (***) satisfies ideal conditions, the BLUE must be OLS on (***)). But,

$$(X'V'VX)^{-1}X'V'Vy = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y = \tilde{\beta}.$$ 

Comment:
$\tilde{\beta}$ is unbiased (consistent) and BLUE. It is also efficient (asymptotically efficient) if $\varepsilon$ is normal.

Theorem:

$$Cov(\tilde{\beta}) = \sigma^2 (X'\Omega^{-1}X)^{-1}.$$ 

Proof:

$$\tilde{\beta} = \beta + (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\varepsilon.$$ 

$$Cov(\tilde{\beta}) = E\left( (\tilde{\beta} - \beta)(\tilde{\beta} - \beta)' \right)$$

$$= E \left( (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\varepsilon\Omega^{-1}X(X'\Omega^{-1}X)^{-1} \right)$$

$$= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}E(\varepsilon\varepsilon')\Omega^{-1}X(X'\Omega^{-1}X)^{-1}$$

$$= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\sigma^2\Omega^{-1}X(X'\Omega^{-1}X)^{-1}$$

$$= \sigma^2 (X'\Omega^{-1}X)^{-1}.$$ 

GLS-3
**Theorem:**

\( \tilde{\beta} \) is efficient relative to \( \hat{\beta} \).

**Proof:**

- \( \text{Cov}(\hat{\beta}) = \sigma^2 (X'X)^{-1}X'\Omega X(X'X)^{-1}. \)
- \( \text{Cov}(\tilde{\beta}) = \sigma^2 (X'\Omega^{-1}X)^{-1}. \)
- It is enough to show that \( (X'X)^{-1}X'\Omega X(X'X)^{-1} - (X'\Omega^{-1}X)^{-1} \) is positive semidefinite. But showing this is equivalent to showing that \( X'\Omega^{-1}X - (X'X)(X'\Omega X)^{-1}(X'X) \) is positive semidefinite.
- Define \( P = X'\Omega^{-1} - (X'X)(X'\Omega X)^{-1}X' \). Then, it can be shown that
  \[
  X'\Omega^{-1}X - (X'X)(X'\Omega X)^{-1}(X'X) = P\Omega P',
  \]
  which is positive semidefinite.

**Theorem:**

Let \( \tilde{\varepsilon} \) be the residual vector from OLS on \(Vy = VX\beta + V\varepsilon \). Then,

\[ \tilde{\sigma}^2 = \tilde{\varepsilon}'\tilde{\varepsilon}/(T - k) \]

is an unbiased and consistent estimator of \( \sigma^2 \).

**Proof:**

Note that \(Vy = VX\beta + V\varepsilon \) satisfies ideal conditions. Therefore, the unbiased and efficient estimator of \( \sigma^2 \) is given by \( s^2 \) from OLS on \(Vy = VX\beta + V\varepsilon \). That is,

\[ \text{SSE }/(T - k) = (Vy - VX\tilde{\beta})'(Vy - VX\tilde{\beta})/(T - k) = \tilde{\varepsilon}'\tilde{\varepsilon }/(T - k). \]
Note:
1) All usual tests can be done directly to $V y = VX\beta + V\epsilon$.

2) $\tilde{\beta} \sim N(\beta, \sigma^2 (X'\Omega^{-1}X)^{-1})$; $\frac{(T-k)\tilde{\sigma}^2}{\sigma^2} \sim \chi^2(T-k)$;

and $\tilde{\beta}$ and $\tilde{\sigma}^2$ are stochastically independent.

3) Even if $\epsilon$ is not normal, 2) holds if $T$ is large.

(3.2) $\Omega$ is not known

Assumption:

Let $\Omega (T \times T)$ depend on a $p \times 1$ vector, $\theta$ ($p < T$): $\Omega = \Omega(\theta)$.

Examples:

1) AR(1): $\epsilon_t = \rho \epsilon_{t-1} + v_t$, $v_t$ iid with $N(0, \sigma^2)$. $\rightarrow \Omega$ depends on $\rho$.

2) ARCH: Autoregressive Conditional Heteroskedasticity.

2.1) Let $\Omega_{t-1}$ be the set of information available at time $t-1$.

2.2) $\epsilon_t \sim N(0, h_t)$, where $h_t = \text{var}(\epsilon_t|\Omega_{t-1})$ and,

$$h_t = \omega + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2 + .... + \alpha_p \epsilon_{t-p}^2.$$ 

$\rightarrow$ Called ARCH(p) model.

$\rightarrow \Omega$ depends on $\omega$ and $\alpha_1, ... , \alpha_p$.

Theorem:

$\hat{\Omega} = \Omega(\hat{\theta})$ is consistent for $\Omega$ if $\hat{\theta}$ is consistent for $\theta$.

GLS-5
Definition:

A feasible GLS (FGLS) is given by \( \hat{\beta}_f = (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}y \).

Comments:

1) No reason to believe that FGLS and GLS are always asymptotically equivalent even if T is large. [For example, See Schmidt.]
2) But, often if X is nonstochastic.

[4] Efficiency of GLS

- Maximum-Likelihood Estimator (MLE)
- Assume that \( \varepsilon \sim N(0_{T\times1}, \sigma^2\Omega(\theta)) \). Then, log-likelihood function is:
  \[
  l_T(\beta, \sigma^2, \theta) = \text{constant} - (T/2)\ln(\sigma^2) - (1/2)\ln[\det(\Omega(\theta))] \\
  - \{1/(2\sigma^2)\}(y-X\beta)'\Omega(\theta)^{-1}(y-X\beta). 
  \]
- MLE of \( \beta, \sigma^2 \) and \( \theta \) are obtained by maximizing \( l_T(\beta, \sigma^2, \theta) \). These MLEs are efficient when T is large.

Almost Theorem:

\( \tilde{\beta}_f \approx \hat{\beta} \approx \hat{\beta}_{MLE} \), where T is large and X is nonstochastic (strictly exogenous).

[See Schmidt for a counterexample for this almost theorem.]
Comments:

- When \( y = X\beta + \varepsilon \) satisfies SIC other than \( \Omega \neq I_T \), \( Vy = VX\beta + V\varepsilon \) satisfies all of SIC.
- When \( y = X\beta + \varepsilon \) satisfies WIC other than \( \Omega \neq I_T \), \( Vy = VX\beta + V\varepsilon \) might violate WIC. It might be the case that
  \[
p \lim_{T \to \infty} \frac{1}{T} XY'Y\varepsilon = p \lim_{T \to \infty} \frac{1}{T} X'\Omega^{-1}\varepsilon \neq 0_{k \times 1}.
  \]

Definition:

- We say that the regressors \( x_{t*} \) are **weakly exogenous** with respect to the \( \varepsilon_t \) if
  \( E(\varepsilon_t | x_{t*}, x_{t-1*}, \ldots, x_{1*}) = 0 \) for any \( t \).
- We say that the regressors \( x_{t*} \) are **strictly exogenous** with respect to the \( \varepsilon_t \) if
  \( E(\varepsilon_t | x_{T*}, x_{T-1*}, \ldots, x_{1*}) = 0 \) for any \( t \).

- Note that WIC only requires weakly exogenous regressors.
  - For cross-section data, the regressors are most likely to be strictly exogenous. But, strictly exogenous regressors are rare in time-series data models.
  - When regressors are weakly exogenous, GLS may be inconsistent. Even when GLS and FGLS are consistent, the asymptotic distributions of GLS and FGLS can be different for some cases.
- If we strengthen WIC with the assumption of strictly exogenous regressors, \( Vy = VX\beta + V\varepsilon \) satisfies WIC.