Advanced Econometrics

Chapter 8: Generalized Least Squares Estimation

In Choi

Sogang University

The model and assumptions

- Model

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

- Assumptions on the error terms

$$(i)\ E\left[\varepsilon|X\right] = 0$$

$$(ii)\ E\left[\varepsilon\varepsilon'|X\right] = \begin{cases} \sigma^2\Omega \\ \Sigma \end{cases} \quad (\Omega,\ \Sigma > 0)\,.$$

The model and assumptions

1. Heteroskedasticity

$$\Sigma = \begin{bmatrix} \sigma_1^2 & & & \underset{\sim}{0} \\ & \ddots & & \\ & & \ddots & \\ \underset{\sim}{0} & & & \sigma_n^2 \end{bmatrix}$$

2. Autocorrelation

$$\sigma^2\Omega = \sigma^2 \begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{n-1} \\ \rho_1 & 1 & \cdots & \rho_{n-2} \\ \vdots & & \ddots & \\ \rho_{n-1} & \cdots & \cdots & 1 \end{bmatrix}$$

The model and assumptions

Example

Let

$$\varepsilon_t = v_t + \theta v_{t-1}, \; v_t \sim iid(0, \sigma_v^2).$$

Then,

$$\begin{aligned} Var(\varepsilon_t) &= \sigma^2(1+\theta^2); \\ Cov(\varepsilon_t \varepsilon_{t\pm 1}) &= \sigma^2\theta; \\ Cov(\varepsilon_t \varepsilon_{t\pm k}) &= 0 \text{ for } k \geq 2. \end{aligned}$$

Thus

$$\sigma^2\Omega = \sigma^2 \begin{bmatrix} 1+\theta^2 & \theta & & \underset{\sim}{0} \\ \theta & 1+\theta^2 & \ddots & \\ & \ddots & \ddots & \theta \\ \underset{\sim}{0} & & \theta & 1+\theta^2 \end{bmatrix}$$

- Since $\Sigma > 0$, it can be factored as

$$\Sigma = C\Lambda C'$$

where the columns of C are the characteristic vectors of Σ and the characteristic roots of Σ are put in the diagonal matrix Λ.

GLS

Example

Let

$$\Sigma = \begin{bmatrix} 1 & 0.5 & 0.25 \\ 0.5 & 1 & 0.5 \\ 0.25 & 0.5 & 1 \end{bmatrix}.$$

Then

$$\begin{aligned} \Lambda &= diag(0.41, 0.75, 1.84); \\ C &= \begin{bmatrix} 0.45 & -0.71 & 0.54 \\ -0.77 & 0.00 & 0.64 \\ 0.45 & 0.71 & 0.54 \end{bmatrix}. \end{aligned}$$

The columns of C contain the eigenvectors of Σ. The eigenvectors are orthonormal.

- Let $P' = C\Lambda^{-1/2}$. Then

$$\begin{aligned}\Sigma^{-1} &= C'^{-1}\Lambda^{-1}C^{-1} = C\Lambda^{-1}C' = C\Lambda^{-1/2}\Lambda^{-1/2}C' \\ &= P'P\end{aligned}$$

since $C' = C^{-1}$ (i.e., C is an orthogonal matrix). Premultiplying the linear regression model by P, we obtain

$$Py = PX\beta + P\varepsilon$$

or

$$y_* = X_*\beta + \varepsilon_*.$$

GLS

- Hence

$$\begin{aligned} E\left(\varepsilon_* \varepsilon_*'\right) &= PE\left(\varepsilon\varepsilon'\right) P' = P\Sigma P' \\ &= \Lambda^{-1/2} C' C \Lambda C' C \Lambda^{-1/2} \\ &= I. \end{aligned}$$

The transformed model satisfies the conditions of the classical linear regression model. Hence

$$\begin{aligned} \hat{\beta}_{GLS} &= \left(X_*' X_*\right)^{-1} X_*' y_* \\ &= \left(X' P' P X\right)^{-1} X' P' P y \\ &= \left(X' \Sigma^{-1} X\right)^{-1} X' \Sigma^{-1} y. \end{aligned}$$

This estimator is called the generalized least squares estimator.

GLS

- Properties of the GLS estimator

1. If $E\left[\varepsilon_* | X_*\right] = 0$, $E\left[\hat{\beta}_{GLS}\right] = \beta$.
2. If $\frac{1}{n} X_*' X_* \xrightarrow{p} Q_* \left(> 0\right)$, $\hat{\beta}_{GLS} \xrightarrow{p} \beta$.
3. $\sqrt{n}\left(\hat{\beta}_{GLS} - \beta\right) \xrightarrow{d} N\left(0, Q_*^{-1}\right)$.

- The GLS estimator $\hat{\beta}_{GLS}$ is the BLUE.

$$y_t = \beta' x_t + \varepsilon_t$$

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t,\ \ u_t \sim iid\left(0, \sigma^2\right),\ \ |\rho| < 1.$$

$$\begin{aligned} E\varepsilon\varepsilon' &= \frac{\sigma^2}{1-\rho^2}\begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & & \cdots & \rho^{n-2} \\ \vdots & & \ddots & & \\ \vdots & & & \ddots & \\ \rho^{n-1} & & & & 1 \end{bmatrix} \\ &= \Sigma. \end{aligned}$$

$$\Sigma^{-1} = \frac{1}{\sigma^2} \begin{bmatrix} 1 & -\rho & & & \underset{\sim}{0} \\ -\rho & 1+\rho^2 & -\rho & & \\ & \ddots & \ddots & \ddots & \\ & & -\rho & 1+\rho^2 & -\rho \\ \underset{\sim}{0} & & & -\rho & 1 \end{bmatrix}.$$

- The transformation matrix is

$$P = \begin{bmatrix} \sqrt{1-\rho^2} & & & \underset{\sim}{0} \\ -\rho & 1 & & \\ & \ddots & \ddots & \\ \underset{\sim}{0} & & -\rho & 1 \end{bmatrix}$$

- The transformed model is

$$\begin{aligned} \sqrt{1-\rho^2} y_1 &= \sqrt{1-\rho^2} x_1' \beta + \varepsilon_1^* \\ y_t - \rho y_{t-1} &= \left(x_t - \rho x_{t-1} \right)' \beta + u_t \end{aligned}$$

where $\varepsilon_1^* = \sqrt{1-\rho^2}\varepsilon_1$. Since
(i) $E\varepsilon_1^* = Eu_t = 0$
(ii) $Var\left(\varepsilon_1^*\right) = \left(1-\rho^2\right) Var\left(\varepsilon_1\right) = \left(1-\rho^2\right) \times \frac{\sigma^2}{1-\rho^2} = \sigma^2$
(iii) $E\left(\varepsilon_1^* u_t\right) = \sqrt{1-\rho^2} E\left(\varepsilon_1 u_t\right) = 0,\ t = 2, \cdots, n,$
The error terms of the transformed model satisfy the condition of the standard linear regression model.

Note When

$$y_t - \rho y_{t-1} = (x_t - \rho x_{t-1})' \beta + u_t$$

used ignoring the first observation, it is called the Cochran-Orcutt procedure.

Feasible GLS

- The GLS estimator $\hat{\beta}_{GLS}$ depends on the unknown parameters associated with Σ and, therefore, cannot be used in practice. Suppose that $\hat{\Sigma} \xrightarrow{p} \Sigma$. Then the feasible GLS estimator is defined by

$$\hat{\beta}_{FGLS} = \left(X'\hat{\Sigma}^{-1}X\right)^{-1} X'\hat{\Sigma}^{-1}y.$$

If

$$\frac{1}{n}X'\hat{\Sigma}^{-1}X - \frac{1}{n}X'\Sigma^{-1}X \xrightarrow{p} 0$$

and

$$\frac{1}{\sqrt{n}}X'\hat{\Sigma}^{-1}\varepsilon - \frac{1}{\sqrt{n}}X'\Sigma^{-1}\varepsilon \xrightarrow{p} 0,$$

$\hat{\beta}_{GLS}$ and $\hat{\beta}_{FGLS}$ have the same asymptotic distribution.

Feasible GLS

Example

AR(1) error

$$
\begin{aligned}
y_t &= \beta' x_t + \varepsilon_t \\
\varepsilon_t &= \rho \varepsilon_{t-1} + u_t, \;\; u_t \sim iid\left(0, \sigma^2\right), \;\; |\rho| < 1.
\end{aligned}
$$

(i) Run OLS and get $\hat{\varepsilon}_t$.
(ii) Run $AR(1)$ regression using $\hat{\varepsilon}_t$. This gives $\hat{\rho}$.
(iii) Transform the model using $\hat{\rho}$ and run OLS.

Equivalence of GLS and OLS

- Let $X'X$ and Σ be both positive definite. Then the following statements are equivalent
  - (A) $(X'X)^{-1} X'\Sigma X (X'X)^{-1} = \left(X'\Sigma^{-1}X\right)^{-1}$.
  - (B) $\Sigma X = XB$ for some nonsingular B.
  - (C) $(X'X)^{-1} X' = \left(X'\Sigma^{-1}X\right)^{-1} X'\Sigma^{-1}$.

Example

$$
\begin{aligned}
y_t &= \beta_0 + \beta_1 t + \varepsilon_t \\
\varepsilon_t &= \rho \varepsilon_{t-1} + u_t, \; u_t \sim iid\left(0, \sigma^2\right), \;\; |\rho| < 1.
\end{aligned}
$$

$$
\Sigma = \frac{\sigma^2}{1-\rho^2} \begin{bmatrix} 1 & \rho & \cdots & \rho^{n-1} \\ \rho & 1 & \cdots & \rho^{n-2} \\ & & \ddots & \\ \rho^{n-1} & & & 1 \end{bmatrix}
$$

Then

$$
\left(X'X\right)^{-1} X' \simeq \left(X'\Sigma^{-1}X\right)^{-1} X'\Sigma^{-1} \text{ (This is an exercise problem).}
$$

Thus, OLS and GLS for this model are asymptotically equivalent. (cf. Grenander and Rosenblatt, 1958).

GLS underheteroskedasticity

- Suppose that $Var(\varepsilon_i|x_i) = \sigma_i^2$. The GLS estimator is obtained by regressing

$$Py = \begin{bmatrix} y_1/\sigma_1 \\ \vdots \\ y_n/\sigma_n \end{bmatrix} \text{ on } Px = \begin{bmatrix} x_1/\sigma_1 \\ \vdots \\ x_n/\sigma_n \end{bmatrix}$$

This gives the GLS estimator

$$\hat{\beta}_{GLS} = \left[\sum_{i=1}^{n}\left(\frac{1}{\sigma_i^2}\right)x_i x_i'\right]^{-1}\left[\sum_{i=1}^{n}\left(\frac{1}{\sigma_i^2}\right)x_i y_i\right].$$

GLS

- If $\sigma_i^2 = x_i'\alpha$, we may write

$$\varepsilon_i^2 = \sigma_i^2 + \nu_i$$

where

$$\nu_i = \varepsilon_i^2 - E\left(\varepsilon_i^2 | x_i\right).$$

- Replacing ε_i^2 with e_i^2, we have an approximate relation

$$e_i^2 = x_i'\alpha + \nu_i^*.$$

- Running OLS on this equation, we can obtain $\hat{\alpha}$ and $\hat{\sigma}_i^2 = x_i'\hat{\alpha}$. The feasible GLS estimator is obtained plugging $\hat{\sigma}_i$ into the formula of GLS.
- In practice, $x_i'\hat{\alpha}$ may become negative which is problematic.

GLS

- We may use other models for heteroskedasticity.

Example

$$\begin{aligned} \sigma_i^2 &= (x_i'\alpha)^2 \\ \sigma_i^2 &= \exp(x_i'\alpha) \\ &\vdots \end{aligned}$$

Exercises

1. Consider the model

$$y_t = \mu + \varepsilon_t, (t = 1, ..., n).$$

It is known that $\Sigma^{-1} = diag(1, \lambda, ..., \lambda^{n-1})$, where $\Sigma = E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')$ and $\boldsymbol{\varepsilon} = [\varepsilon_1, ..., \varepsilon_n]'$.

   1. Letting $\lambda = 1$, derive the asymptotic distribution of $\hat{\mu}_{GLS}$.[1]
   2. Derive the asymptotic distribution of the feasible GLS estimator of μ that sues a consistent estimator of λ, $\hat{\lambda} = \lambda + \frac{1}{n}$. Assume $\lambda = 1$.
   3. Do the GLS and feasible GLS estimators have the same distribution in the limit?

2. Consider the model

$$y_t = \beta_0 + \beta_1 t + \varepsilon_t, (t = 1, ..., n)$$

where $\varepsilon_t = \rho\varepsilon_{t-1} + \omega_t, |\rho| < 1, \omega_t \sim iid(0, 1)$. Using Kruskal's theorem, show that OLS and GLS are asymptotically equivalent.

[1] $\lim_{n\to\infty}(1 + \frac{1}{n})^n = e$.