Topic 4: Generalized Least Squares

Generalized Least Squares (Alexander Aitken, 1935)

- In the present context, (Ordinary) LS ignores some important information, and we'd anticipate that this will result in a loss of efficiency when estimating $\boldsymbol{\beta}$.
- Let's see how to obtain the fully efficient (linear unbiased) estimator.
- Recall that $V(\boldsymbol{\varepsilon}) = E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'] = \Sigma = \sigma^2\Omega$.
- Generally, Ω will be *unknown*. However, to begin with, let's consider the case where it is actually *known*.
- Clearly, Ω must be *symmetric*, as it is a covariance matrix.
- Suppose that Ω is also *positive-definite*.
- Then, Ω^{-1} is also positive-definite, and so there exists a *non-singular* matrix, P, such that $\Omega^{-1} = P'P$.
- Our model is:

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad ; \quad \boldsymbol{\varepsilon} \sim [0\,, \sigma^2\Omega]$$

- Pre-multiply the equation by P:

$$P\boldsymbol{y} = PX\boldsymbol{\beta} + P\boldsymbol{\varepsilon}$$

or,

$$\boldsymbol{y}^* = X^*\boldsymbol{\beta} + \boldsymbol{\varepsilon}^* \quad ; \quad \text{say}$$

- Now, Ω is non-random, so P is also non-random.
- So, $\quad E[\boldsymbol{\varepsilon}^*] = E[P\boldsymbol{\varepsilon}] = P\,E[\boldsymbol{\varepsilon}] = \mathbf{0}$.
- And $\quad V[\boldsymbol{\varepsilon}^*] = V[P\boldsymbol{\varepsilon}]$

$$= PV(\boldsymbol{\varepsilon})P'$$

$$= P(\sigma^2\Omega)P' = \sigma^2 P\Omega P'$$

- Note that $\quad P\Omega P' = P(\Omega^{-1})^{-1}P'$

$$= P(P'P)^{-1}P'$$

$$= PP^{-1}(P')^{-1}P' = I$$

- (Because P is both square and non-singular.)

- So, $E[\boldsymbol{\varepsilon}^*] = \mathbf{0}$ and $V[\varepsilon^*] = \sigma^2 I$.
- The transformed model, $\boldsymbol{y}^* = X^*\boldsymbol{\beta} + \boldsymbol{\varepsilon}^*$, has an error-term that satisfies the *usual assumptions*. **In particular, it has a scalar covariance matrix**.

- So, if we apply (Ordinary) Least Squares to the model, $\boldsymbol{y}^* = X^*\boldsymbol{\beta} + \boldsymbol{\varepsilon}^*$, we'll get the BLU estimator of $\boldsymbol{\beta}$, by the Gauss-Markhov Theorem.

- We call this the **Generalized Least Squares Estimator** of $\boldsymbol{\beta}$.

- The formula for this estimator is readily determined:

$$\hat{\boldsymbol{\beta}} = [X^{*\prime}X^*]^{-1}X^{*\prime}\boldsymbol{y}^*$$
$$= [(PX)'(PX)]^{-1}(PX)'(P\boldsymbol{y})$$
$$= [X'P'PX]^{-1}X'P'P\boldsymbol{y}$$
$$= [X'\Omega^{-1}X]^{-1}X'\Omega^{-1}\boldsymbol{y}$$

- Clearly, because $E[\boldsymbol{\varepsilon}^*] = \mathbf{0}$ as long as the regressors are non-random, the GLS estimator, $\hat{\boldsymbol{\beta}}$ is **unbiased**.
- Moreover, the covariance matrix of the GLS estimator is:

$$V(\hat{\boldsymbol{\beta}}) = [X'\Omega^{-1}X]^{-1}X'\Omega^{-1}V(\boldsymbol{y})\{[X'\Omega^{-1}X]^{-1}X'\Omega^{-1}\}'$$
$$= [X'\Omega^{-1}X]^{-1}X'\Omega^{-1}\sigma^2\Omega\Omega^{-1}X[X'\Omega^{-1}X]^{-1}$$
$$= \sigma^2[X'\Omega^{-1}X]^{-1} \, .$$

- If the errors are Normally distributed, then the full sampling distribution of the GLS estimator of $\boldsymbol{\beta}$ is:

$$\hat{\boldsymbol{\beta}} \sim N[\boldsymbol{\beta} \, , \sigma^2[X'\Omega^{-1}X]^{-1}]$$

- The GLS estimator is just the OLS estimator, applied to the transformed model, and the latter model satisfies all of the usual conditions.

- So, the *Gauss-Markhov Theorem* applies to the GLS estimator.
- The GLS estimator is BLU for this more general model (with a non-scalar error covariance matrix).
- Note: OLS must be *inefficient* in the present context.

"Feasible" GLS Estimator

- In order to be able to implement the GLS estimator, in practice, we're usually going to have to provide a *suitable estimator* of Ω (or Σ).
- Presumably we'll want to obtain an estimator that is *at least consistent*, and place this into the formula for the GLS estimator, yielding:

$$\tilde{\boldsymbol{\beta}} = \left[X'\hat{\Omega}^{-1}X\right]^{-1}X'\hat{\Omega}^{-1}\boldsymbol{y}$$

- Problem: The Ω matrix is $(n \times n)$, and it has $n(n+1)/2$ *distinct* elements. However, we have only n observations on the data. This precludes obtaining a consistent estimator.
- We need to constrain the elements of Ω in some way.
- In practice, this won't be a big problem, because usually there will be lots of "structure" on the form of Ω .
- Typically, we'll have $\Omega = \Omega(\boldsymbol{\theta})$, where the vector, $\boldsymbol{\theta}$ has low dimension.

Example: Heteroskedasticity

Suppose that $var.(\varepsilon_i) \propto (\theta_1 + \theta_2 z_i) = \sigma^2(\theta_1 + \theta_2 z_i)$

Then,

$$\Omega = \begin{pmatrix} \theta_1 + \theta_2 z_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \theta_1 + \theta_2 z_n \end{pmatrix}$$

There are just two parameters that have to be estimated, in order to obtain $\hat{\Omega}$.

Example: **Autocorrelation**

Suppose that the errors follow a *first-order autoregressive process*:

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t \quad ; \quad u_t \sim N[0\,, \sigma_u^2] \qquad \text{(i.i.d.)}$$

Then (for reasons we'll see later),

$$\Omega = \frac{\sigma_u^2}{1-\rho^2}\begin{bmatrix} 1 & \rho & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \rho^{n-2} \\ \vdots & \rho & \ddots & \vdots \\ \rho^{n-1} & \cdots & \cdots & 1 \end{bmatrix} = \Omega(\rho).$$

- So, typically, we'll just have to estimate a very small number of parameters in order to get an estimator of Ω .
- As long as we use a *consistent estimator* for these parameters – the elements of θ, this will give us a consistent estimator of Ω and of Ω^{-1}, by Slutsky's Theorem.
- This in turn, will ensure that our Feasible GLS estimator of β is also *weakly consistent*:

$$plim(\tilde{\beta}) = plim\left\{[X'\hat{\Omega}^{-1}X]^{-1}X'\hat{\Omega}^{-1}\boldsymbol{y}\ \right\}$$

$$= plim\{[X'\Omega^{-1}X]^{-1}X'\Omega^{-1}\boldsymbol{y}\ \}$$

$$= plim(\hat{\beta}) = \beta\ .$$

- Also, if $\hat{\Omega}$ is consistent for Ω then $\tilde{\beta}$ will be *asymptotically efficient*.
- In general, we can say little about the *finite-sample* properties of our feasible GLS estimator.
- Usually it will be *biased*, and the nature of the bias will depend on the form of Ω and our choice of $\hat{\Omega}$.
- In order to apply either the GLS estimator, or the feasible GLS estimator, we need to know the form of Ω .

Typically, this is achieved by postulating various forms, and testing to see if these are supported by the data.