AAE 637 Lab 4: Generalized Least Squares Estimator

TA: Charng-Jiun Yu

February 14, 2018

1 Motivation

- When the errors are correlated, OLS estimator is consistent but inefficient. Generalized least squares (GLS) estimator can be more efficient.

- Examples of correlated errors:
  1. Autoregressive errors from AR(m) models:

     $$y_t = \mathbf{x_t'}\boldsymbol{\beta} + u_t$$
     $$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + ... + \rho_m u_{t-m} + v_t$$

     where v_t is i.i.d., $E(v_t) = 0$, and $E(v_t^2) = \sigma_v^2$. This is the case when present errors are correlated with past errors.
  2. Correlated errors within a group: we might believe that errors are correlated within a certain group. For example, when analyzing the impact of job injury risk on wages, the errors of observations in the same industry might be correlated with each other because we might systematically overpredict (or underpredict) wages given an industry.

- Note that heteroskedasticity means the variance of errors varies across observations (unlike homoskedastic errors), but it tells nothing about the correlation between errors. The heteroskedastic-robust standard error estimator we commonly use actually assumes independent errors.

2 Basic Idea

- Suppose we have a linear model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$. For GLS estimators we assume that the conditional variance of error terms is **known**:

  $$Cov(u|\mathbf{X}) = E(\mathbf{uu'}|\mathbf{X}) = \sigma^2\boldsymbol{\Omega}$$

  where $\boldsymbol{\Omega}$ is a nonsingular matrix and $Cov(u|\mathbf{X})$ is the variance-covariance matrix of error terms.

  GLS estimator minimizes the weighted sum of squared errors:

  $$WSSE = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Omega}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

We have an analytical solution for GLS estimator in this linear model:

$$\boldsymbol{\beta}_{GLS} = (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y})$$

The asymptotic variance is

$$Var(\boldsymbol{\beta}_{GLS}) = \sigma^2(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}$$

- Properties of GLS estimator: unbiased, consistent, efficient, and asymptotic normal.
- GLS is equivalent to using OLS on a linearly transformed data. That is, there exists a matrix $\mathbf{P}$ such that $\boldsymbol{\Omega}^{-1} = \mathbf{P}'\mathbf{P}$, $\mathbf{X}^* = \mathbf{PX}$, and $\mathbf{y}^* = \mathbf{Py}$. We can express the GLS estimator as

$$\boldsymbol{\beta}_{GLS} = (\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}(\mathbf{X}^{*\prime}\mathbf{y}^*)$$

3 Feasible GLS

- In reality, the conditional variance of error terms, $\sigma^2\boldsymbol{\Omega}$, is **unknown**. We have to estimate it. The resulting estimator is the feasible GLS (FGLS) estimator:

$$\hat{\boldsymbol{\beta}}_{FGLS} = (\mathbf{X}'\hat{\boldsymbol{\Omega}}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\hat{\boldsymbol{\Omega}}^{-1}\mathbf{y})$$

and the asymptotic variance:

$$Var(\hat{\boldsymbol{\beta}}_{FGLS}) = \hat{\sigma}^2(\mathbf{X}'\hat{\boldsymbol{\Omega}}^{\mathbf{-1}}\mathbf{X})^{-1}$$

- Properties of FGLS are identical to GLS **when the sample size is large**, given that $\boldsymbol{\Omega}$ is correctly specified and consistently estimated. In small sample, FGLS is not necessarily more efficient than OLS.

- There are many ways of estimating $\boldsymbol{\Omega}$. Here is the example for an AR(1) model:

$$y_t = \mathbf{x_t}'\boldsymbol{\beta} + u_t$$

$$u_t = \rho u_{t-1} + v_t$$

$$t = 1, ..., T$$

where v_t is i.i.d., $E(v_t) = 0$, and $E(v_t^2) = \sigma_v^2$. We also assume a stationary process $|\rho| < 1$.

The variance-covariance matrix of errors can be expressed as:

$$E(\mathbf{uu}') = \begin{pmatrix} Var(u_1) & Cov(u_2, u_1) & \cdots & Cov(u_T, u_1) \\ Cov(u_1, u_2) & Var(u_2) & \cdots & Cov(u_T, u_2) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(u_1, u_T) & Cov(u_2, u_T) & \cdots & Var(u_T) \end{pmatrix}$$

Assume homoskadastic errors for u and v, we have the diagonal terms

$$\begin{aligned} & Var(u_t) = \rho^2 Var(u_{t-1}) + Var(v_t) \\ \Rightarrow \quad & \sigma_u^2 = \rho^2\sigma_u^2 + \sigma_v^2 \\ \Rightarrow \quad & \sigma_u^2 = \frac{\sigma_v^2}{1-\rho^2} \end{aligned}$$

For the covariances, note that $cov(u_t, u_{t'}) = E(u_t u_{t'})$. To start simple, let's look at the covariance of u_t and u_{t-1}:

$$E[u_t u_{t-1}] = E[(\rho u_{t-1} + v_t)u_{t-1}] = \rho E(u_{t-1}^2) + E(v_t u_{t-1}) = \rho\sigma_u^2 + E(v_t u_{t-1})$$

What is $E(v_t u_{t-1})$? Intuitively it should be zero, because the error u in the past should not be affected by its regression error v in the future. We can also find out from the error structure.

Substitute u_{t-1} in the error equation we can get

$$u_t = \rho(\rho u_{t-2} + v_{t-1}) + v_t = \rho^2 u_{t-2} + \rho v_{t-1} + v_t$$

If we continue this process (by substituting u_{t-2} and so on), u_t can be expressed as

$$u_t = \rho^{T-1}u_1 + \rho^{T-2}v_2 + ... + \rho v_{t-1} + v_t = \sum_{i=0}^{T-2} \rho^i v_{t-i} + \rho^{T-1}u_1$$

Note that none of the v_{t-i} on the right-hand side is after period t. This implies that u is not correlated with v in the future periods.

As $E(v_t u_{t-1}) = 0$, We have

$$E(u_t u_{t-1}) = \rho\sigma_u^2 + E(v_t u_{t-1}) = \rho\sigma_u^2$$

Let's try the covariance of u_t and u_{t-2}:

$$E(u_t u_{t-2}) = E[(\rho u_{t-1}+v_t)u_{t-2}] = \rho E(u_{t-1}u_{t-2})+E(v_t u_{t-2}) = \rho E[(\rho u_{t-2}+v_{t-1})u_{t-2}] = \rho^2 E(u_{t-2}) = \rho^2\sigma_u^2$$

Following this logic, we can write the general form for these off-diagonal terms as

$$E(u_t u_{t-i}) = \rho^i \sigma_u^2 = \rho^i \frac{\sigma_v^2}{1-\rho^2}$$

The last equality comes from our previous result $\sigma_u^2 = \frac{\sigma_v^2}{1-\rho^2}$.

Now we have our variance-covariance matrix for u:

$$E(\mathbf{uu}') = \begin{pmatrix} \frac{\sigma_v^2}{1-\rho^2} & \frac{\rho\sigma_v^2}{1-\rho^2} & \cdots & \frac{\rho^{T-1}\sigma_v^2}{1-\rho^2} \\ \frac{\rho\sigma_v^2}{1-\rho^2} & \frac{\sigma_v^2}{1-\rho^2} & \cdots & \frac{\rho^{T-2}\sigma_v^2}{1-\rho^2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\rho^{T-1}\sigma_v^2}{1-\rho^2} & \frac{\rho^{T-2}\sigma_v^2}{1-\rho^2} & \cdots & \frac{\sigma_v^2}{1-\rho^2} \end{pmatrix} = \sigma_v^2 \frac{1}{1-\rho^2} \begin{pmatrix} 1 & \rho & \cdots & \rho^{T-1} \\ \rho & 1 & \cdots & \rho^{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \cdots & 1 \end{pmatrix} = \sigma_v^2 \mathbf{\Omega}$$

We can estimate $\mathbf{\Omega}$ by

$$\hat{\mathbf{\Omega}} = \frac{1}{1-\hat\rho^2} \begin{pmatrix} 1 & \hat\rho & \cdots & \hat\rho^{T-1} \\ \hat\rho & 1 & \cdots & \hat\rho^{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \hat\rho^{T-1} & \hat\rho^{T-2} & \cdots & 1 \end{pmatrix}$$

- We still need $\hat\sigma_v^2$ to get the conditional variance of error terms. Similar to the case in OLS, $\hat\sigma_v^2$ can be obtained from the weighted sum of squared residuals:

$$\hat\sigma_v^2 = \frac{(\mathbf{y}-\mathbf{X}\hat{\boldsymbol\beta}_{FGLS})'\hat{\mathbf{\Omega}}^{-1}(\mathbf{y}-\mathbf{X}\hat{\boldsymbol\beta}_{FGLS})}{T-k}$$

where k is the number of parameters in $\hat{\boldsymbol\beta}_{FGLS}$.

- Let's sum up the two-step procedure of estimating FGLS estimators using the above example:
  1. From $y_t = \mathbf{x_t}'\boldsymbol{\beta} + u_t$, estimate β using the OLS estimator and get the residual $\hat{u}_t$
  2. From $u_t = \rho u_{t-1} + v_t$, estimate ρ using the OLS estimator by replacing u_t with $\hat{u}_t$ (also u_{t-1} with $\hat{u}_{t-1}$) obtained from the previous step
  3. Given $\hat{\rho}$, get $\hat{\boldsymbol{\Omega}}$
  4. Given $\hat{\boldsymbol{\Omega}}$, obtain $\hat{\boldsymbol{\beta}}_{FGLS}$
  5. Using both $\hat{\boldsymbol{\Omega}}$ and $\hat{\boldsymbol{\beta}}_{FGLS}$, we can get $\hat{\sigma}_v^2$ and therefore $Var(\hat{\boldsymbol{\beta}}_{FGLS})$

4 Hypothesis test on AR(1) error structure

- The simplest method is a t-test on $\rho = 0$. From $\boldsymbol{\Omega}$, we can see that if $\rho = 0$, the autocorrelation does not exist. Therefore, we can test the existence of AR(1) structure with the null hypothesis $H_0 : \rho = 0$ using t-test.

- We can also do the Durbin-Watson test that can be more accurate than the above t-test in finite samples. The Durbin-Watson statistic is given by:

$$DW = \frac{\sum_{t=2}^{T}(\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{T}\hat{u}_t^2}$$

  This statistic is approximately $2(1-\rho)$. Therefore, DW will be closed to 2 if ρ is closed to 0. We can also see that a positive autocorrelation $0 < \rho < 1$ implies $DW < 2$ and negative autocorrelation $-1 < \rho < 0$ implies $DW > 2$. DW must lie between 0 and 4.

  Here is an example of testing the existence of positive autocorrelation:

$$H_0 : \rho = 0, \quad H_1 : \rho > 0$$

  We reject H_0 if DW is smaller than its lower critical value, and do not reject H_0 when DW is larger than its upper critical value. If DW is between these two critical values, our test is inconclusive. Note that we cannot use Durbin-Watson test if our first-step regression, $y_t = \mathbf{x_t}'\boldsymbol{\beta} + u_t$, has lagged terms.

5 Summary of Common Error Assumptions

1. Homoskedastic independent errors: homoskedastic independent errors are uncorrelated with each other and the variance is constant across observations.

$$E(\mathbf{uu}'|\mathbf{X}) = \sigma^2\mathbf{I}$$

   where $\mathbf{I}$ is the identity matrix.

2. Heteroskedastic independent errors: heteroskedastic independent errors are uncorrelated with each other but the variance varies across observations.

$$E(\mathbf{uu}'|\mathbf{X}) = \sigma_i^2\mathbf{I}$$

3. Autocorrelated errors: autocorrelated errors are correlated with their past errors.

$$E(\mathbf{u}\mathbf{u}'|\mathbf{X}) = \sigma^2\boldsymbol{\Omega}$$

where Ω is a full, positive definite matrix with 1 on the diagonal. In the case of AR(1),

$$\boldsymbol{\Omega} = \frac{1}{1-\rho^2}\begin{pmatrix} 1 & \rho & \cdots & \rho^{T-1} \\ \rho & 1 & \cdots & \rho^{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \cdots & 1 \end{pmatrix}$$

Note that we assume homoskedasticity. This is common for autocorrelated errors.

4. Clustered errors: clustered errors are correlated with each other in the same cluster (group) but uncorrelated with errors in other clusters.

$$E(u_{ig}u_{jg^*}|x_{ig}, x_{jg^*}) = 0, \text{ unless } g = g^*$$

⋆ Notes on notations: lower-case normal letters (x) denote scalars, lower case boldfaced letters (**x**) denote column vectors, and upper-case boldfaced letters (**X**) denote matrices.