GLS and FGLS

Econ 671

Purdue University

- In this lecture we continue to discuss properties associated with the GLS estimator.
- In addition we discuss the practical issue of what to do when the covariance matrix is unknown (which, of course, it is), leading to a discussion of the *Feasible Generalized Least Squares (FGLS)* estimator.
- We close the lecture by providing a series of models where FGLS can be implemented, leaving most details of the implementation to other courses.

We begin our GLS investigation by deriving the (finite sample) covariance matrix. To this end, we note:

-

Recall that the variance of the OLS estimator in the presence of a general Ω was:

-

Aitken's theorem tells us that the GLS variance is "smaller." This is obvious, right?

To establish this result, note:

-

We claim that this difference can be written in the form:

-

where

-

The proof is simple - just multiply the expression out. $\text{Var}(\hat{\beta}_{OLS}|X)$ is produced directly in the product, while the remaining terms combine to give $\text{Var}(\hat{\beta}_{GLS}|X)$.

The matrix $A\Omega A'$ is clearly positive definite since Ω is positive definite.

GLS Asymptotics

We do not discuss the asymptotic derivations of the GLS estimator here. Essentially, the methods applied for the OLS case can again be applied upon transforming the data. In this regard, we assume

-

where V is finite and nonsingular. Asymptotic normality follows using similar arguments:

-

Feasible Generalized Least Squares

The assumption that Ω is known is, of course, a completely unrealistic one.

In many situations (see the examples that follow), we either suppose, or the model naturally suggests, that Ω is comprised of a finite set of parameters, say α, and once α is known, Ω is also known.

Suppose that a consistent estimator of α is available, and denote this estimator as $\hat{\alpha}$. We could then replace Ω with $\hat{\Omega} = \Omega(\hat{\alpha})$ and implement the *Feasible Generalized Least Squares Estimator* (FGLS):

-

FGLS Asymptotics

To work out the asymptotics of the FGLS estimator, a standard approach is to first demonstrate that FGLS is asymptotically equivalent to GLS, so that the distribution theory for GLS can be "borrowed" and applied to the FGLS estimator (which remains true, asymptotically).

We say that two estimators are asymptotically equivalent if:

-

Sufficient conditions for this to be true are:

-

and

-

A few things to note:

- The preference for FGLS over OLS is an *asymptotic* one. In fact, you can manufacture cases where the OLS estimator is preferable to FGLS in finite samples.
- Interestingly note that FGLS is asymptotically efficient (among the class of linear unbiased estimators) even though we only require a *consistent* estimator of Ω, not necessarily an *efficient* one.
- The finite sample properties of FGLS are quite difficult to work out, in general, as we "use the data twice." Rather surprisingly, however, the FGLS estimator is often unbiased.

FGLS Example #1

Consider the regression model:

-

Here, D_i is an observed, binary dummy variable. Think about this as representing two groups: for example, variances may potentially differ across men and women.

How would estimation proceed here?

To implement FGLS, we require consistent estimates of the variance parameters. How could we obtain such estimates?

-

FGLS Example #1

Thus, in practice, we might proceed as follows:

1. Run OLS and obtain $\hat{\beta}$.

2. Given $\hat{\beta}$, estimate σ_1^2 and σ_2^2 using the formula provided on the last slide.

3. Calculate $\hat{\beta}_{FGLS}$ as:

$$\hat{\beta}_{FGLS} = (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}y, \quad \hat{\Omega} \equiv diag\{D_i\hat{\sigma}_1^2 + (1-D_i)\hat{\sigma}_2^2\}_{i=1}^n.$$

4. Calculate $\widehat{\text{Var}(\hat{\beta}_{FGLS}}|X)$ as:

$$(X'\hat{\Omega}^{-1}X)^{-1}$$

FGLS Example #2: General Parametric Heteroscedasticity

Consider the regression model:

-

- The z_i are observed in the data set. Potentially these are the same as x_i, though z_i may contain more (or fewer) variables than are included in x_i. The first element of z_i is assumed to be an intercept.

- In your study, it is possible that the main interest lies in the conditional *variability* of y rather than (just) the mean. For example, it might be interesting to see how the variability of earnings is related to things like education, test scores, etc. This motivates the adoption of a parametric model to describe the heteroscedasticity. (In contrast to the White approach, where these are merely nuisance parameters).

FGLS Example #2: General Parametric Heteroscedasticity

How would we proceed here?

Again, we recognize that $\hat{\beta}$ is consistent for β and, likewise $\hat{\epsilon}_i^2 \xrightarrow{p} \epsilon_i^2$.
To fix ideas (though this is not necessary), suppose

-

Then,

-

Thus,

-

FGLS Example #2: General Parametric Heteroscedasticity

This suggests that we can estimate α from an OLS regression of $\log(\hat{\epsilon}_i^2)$ on z_i. (Another useful alternative is nonlinear least-squares, though we do not discuss that here).

The intercept parameter included in α, however, will be biased and inconsistent. This relates to the fact that, although the random variable ν_i will have mean 1, the mean of $u_i = \log \nu_i$ is not necessarily zero. In fact, when the errors are normally distributed, the mean of the log of the chi-square is approximately -1.27.

Thus, if we are willing to make the normality assumption, we can "fix" this problem by simply adding 1.27 to the estimated intercept parameter. (**But, does this really matter for purposes of point estimation?**)

FGLS Example #3: Random Effects Panel Models

Consider the regression model:

-

- We now have 2 subscripts: i (for individual or unit) and t for time.
- The primary problem in these models is often to account for correlation in outcomes for a particular unit over time.
- The presence of the random effects α_i accomplishes just that.

FGLS Example #3: Random Effects Panel Models

$$y_{it} = x_{it}\beta + \alpha_i + \epsilon_{it}, \qquad i = 1, 2, \ldots, n, \quad t = 1, 2, \ldots, T.$$

To this we additionally assume that

$$\begin{aligned}
E(\alpha_i | X) &= 0 \\
E(\epsilon_{it} | X) &= 0 \\
\text{Var}(\alpha_i | X) &= \sigma_\alpha^2 \\
\text{Var}(\epsilon_{it} | X) &= \sigma_\epsilon^2 \\
E(\alpha_i \alpha_j | X) &= 0, \quad j \neq i \\
E(\epsilon_{it}\epsilon_{js} | X) &= 0 \quad \text{whenever } i \neq j \text{ or } t \neq s \\
E(\alpha_i \epsilon_{jt} | X) &= 0 \ \forall i, j, t
\end{aligned}$$

FGLS Example #3: Random Effects Panel Models

Now, stack observations for individual i over t. Let

$$y_i = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{bmatrix}$$

What is Var$(y_i|X)$ under the given assumptions?

-

It follows that the entire dependent variable vector, y, stacked across all individuals i in a similar manner has conditional covariance matrix:

-

FGLS Example #3: Random Effects Panel Models

So the panel model, like those before it, fits into the general framework of GLS estimation, where the covariance matrix depends only on a finite number of parameters (in this case, just 2 parameters).

Methods for estimating σ_α^2 and σ_ϵ^2 exist, but are discussed in detail in 672.

FGLS Example #4: Time Series

Consider the regression model:

-

(the variable y_{t-1} is called a *lag* of y_t). Note, then, by substitution:

-

and continuing,

-

so that

-

provided

-

FGLS Example #4: Time Series

Likewise,

-

In a similar manner, we can show:

-

so that

-

Again, this fits within the GLS structure.

FGLS Example #5: SUR Models

Suppose that the outcome for one unit at time t (perhaps the stock return for company 1 at time t) is:

-

Likewise, the return for a second unit is given as:

-

Should we run separate regressions to estimate β_1 and β_2?

FGLS Example #5: SUR Models

-

In particular, there may be *contemporaneous correlation* among the errors that could lead to (asymptotic) efficiency gains via the use of FGLS. That is, suppose:

$$\left[\begin{array}{c}\epsilon_{1t}\\ \epsilon_{2t}\end{array}\right] \stackrel{iid}{\sim} \mathcal{N}\left[\left(\begin{array}{c}0\\0\end{array}\right), \left(\begin{array}{cc}\sigma_1^2 & \sigma_{12}\\ \sigma_{12} & \sigma_2^2\end{array}\right)\right] \equiv \mathcal{N}(0, \Sigma).$$

FGLS Example #5: SUR Models

In this case, we could stack the observations into one "big" regression model, where the resulting covariance matrix will not be diagonal with a constant variance. Specifically, we can write:

-

or

-

where

-

Thus, if we can consistently estimate the three parameters of Σ, there will be asymptotic efficiency gains to using FGLS instead of running separate regressions, equation-by-equation.