Generalized Least Squares

Consider the linear regression model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}, \quad \mathrm{E}(\boldsymbol{u}\boldsymbol{u}^\top) = \boldsymbol{\Omega}, \tag{1}$$

where $\boldsymbol{\Omega}$ is a positive definite $N \times N$ matrix which may depend on $\boldsymbol{X}$.

Suppose we know $\boldsymbol{\Omega}$, or at least can estimate it consistently.

To obtain an efficient estimator of $\boldsymbol{\beta}$, we transform the model so that it satisfies the conditions of the Gauss-Markov theorem.

Estimating the transformed model by OLS yields efficient estimates.

The transformation uses an $N \times N$ matrix $\boldsymbol{\Psi}$, which is usually triangular, that satisfies the equation

$$\boldsymbol{\Omega}^{-1} = \boldsymbol{\Psi}\boldsymbol{\Psi}^\top. \tag{2}$$

For N not too large, we can use a Cholesky decomposition.

The transformed model is

$$\boldsymbol{\Psi}^\top \boldsymbol{y} = \boldsymbol{\Psi}^\top \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\Psi}^\top \boldsymbol{u}. \tag{3}$$

Since $\boldsymbol{\Omega}$ is assumed to be nonsingular, so is $\boldsymbol{\Psi}$.

The OLS estimator of $\boldsymbol{\beta}$ from (3) is

$$\hat{\boldsymbol{\beta}}_{\text{GLS}} = (\boldsymbol{X}^\top \boldsymbol{\Psi}\boldsymbol{\Psi}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{\Psi}\boldsymbol{\Psi}^\top \boldsymbol{y} = (\boldsymbol{X}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{y}. \tag{4}$$

This estimator is called **generalized least squares**, or **GLS**.

The covariance matrix of $\boldsymbol{\Psi}^\top \boldsymbol{u}$ is the identity matrix:

$$\begin{aligned} \mathrm{E}(\boldsymbol{\Psi}^\top \boldsymbol{u}\boldsymbol{u}^\top \boldsymbol{\Psi}) &= \boldsymbol{\Psi}^\top \mathrm{E}(\boldsymbol{u}\boldsymbol{u}^\top)\boldsymbol{\Psi} = \boldsymbol{\Psi}^\top \boldsymbol{\Omega}\boldsymbol{\Psi} \\ &= \boldsymbol{\Psi}^\top (\boldsymbol{\Psi}\boldsymbol{\Psi}^\top)^{-1} \boldsymbol{\Psi} = \boldsymbol{\Psi}^\top (\boldsymbol{\Psi}^\top)^{-1} \boldsymbol{\Psi}^{-1} \boldsymbol{\Psi} = \mathbf{I}. \end{aligned} \tag{5}$$

Thus regression (3) has i.i.d. disturbances. When $\boldsymbol{X}$ is exogenous, the Gauss-Markov theorem applies to it.

If we replace $\boldsymbol{X}$ by $\boldsymbol{\Psi}^\top \boldsymbol{X}$ and σ_0^2 by 1 in the usual formula for the OLS covariance matrix, we find that

$$\text{Var}(\hat{\boldsymbol{\beta}}_{\text{GLS}}) = (\boldsymbol{X}^\top \boldsymbol{\Psi} \boldsymbol{\Psi}^\top \boldsymbol{X})^{-1} = (\boldsymbol{X}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{X})^{-1}. \tag{6}$$

The GLS estimator $\hat{\boldsymbol{\beta}}_{\text{GLS}}$ can also be obtained by minimizing the **GLS criterion function**

$$(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^\top \boldsymbol{\Omega}^{-1} (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}), \tag{7}$$

which is just the SSR from the transformed regression (3).

The GLS estimator $\hat{\boldsymbol{\beta}}_{\text{GLS}}$ is also the solution of the moment conditions

$$\boldsymbol{X}^\top \boldsymbol{\Omega}^{-1} (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{\text{GLS}}) = \boldsymbol{0}. \tag{8}$$

These moment conditions are equivalent to the first-order conditions for the minimization of the GLS criterion function (7).

Equations (8) are a special case of $\boldsymbol{W}^\top (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_W) = \boldsymbol{0}$.

Solving these moment conditions yields the MM estimator

$$\hat{\boldsymbol{\beta}}_W = (\boldsymbol{W}^\top \boldsymbol{X})^{-1} \boldsymbol{W}^\top \boldsymbol{y}. \tag{9}$$

The GLS estimator (4) is a special case of this one, with $\boldsymbol{W} = \boldsymbol{\Omega}^{-1}\boldsymbol{X}$.

In general, the covariance matrix of $\hat{\boldsymbol{\beta}}_W$ is

$$\begin{aligned} \mathrm{Var}(\hat{\boldsymbol{\beta}}_W) &= \mathrm{E}\big((\hat{\boldsymbol{\beta}}_W - \boldsymbol{\beta}_0)(\hat{\boldsymbol{\beta}}_W - \boldsymbol{\beta}_0)^\top\big) && (10)\\ &= \mathrm{E}\big((\boldsymbol{W}^\top \boldsymbol{X})^{-1} \boldsymbol{W}^\top \boldsymbol{u}\boldsymbol{u}^\top \boldsymbol{W} (\boldsymbol{X}^\top \boldsymbol{W})^{-1}\big) && (11)\\ &= (\boldsymbol{W}^\top \boldsymbol{X})^{-1} \boldsymbol{W}^\top \boldsymbol{\Omega}\, \boldsymbol{W} (\boldsymbol{X}^\top \boldsymbol{W})^{-1}. && (12) \end{aligned}$$

For simplicity, we have treated $\boldsymbol{X}$ and $\boldsymbol{W}$ as exogenous here.

The GLS estimator will be at least as efficient as the MM estimator if

$$\boldsymbol{X}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{X} - \boldsymbol{X}^\top \boldsymbol{W} (\boldsymbol{W}^\top \boldsymbol{\Omega}\, \boldsymbol{W})^{-1} \boldsymbol{W}^\top \boldsymbol{X} \tag{13}$$

is positive semidefinite. This can be proved easily enough.

Computing GLS Estimates

Suppose that $\boldsymbol{\Omega} = \sigma^2 \boldsymbol{\Delta}$. We can get away with just knowing $\boldsymbol{\Delta}$, because

$$\hat{\boldsymbol{\beta}}_{\text{GLS}} = (\boldsymbol{X}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{y} = (\boldsymbol{X}^\top \boldsymbol{\Delta}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{\Delta}^{-1} \boldsymbol{y}. \tag{14}$$

The factors of σ^2 and $1/\sigma^2$ cancel out.

If we know $\boldsymbol{\Delta}$ but not $\boldsymbol{\Omega}$, we need to use

$$\widehat{\text{Var}}(\hat{\boldsymbol{\beta}}_{\text{GLS}}) = s^2 (\boldsymbol{X}^\top \boldsymbol{\Delta}^{-1} \boldsymbol{X})^{-1}, \tag{15}$$

where s^2 is the usual OLS estimate of the error variance from (3).

The standard formula $(\boldsymbol{X}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{y}$ requires either the inverse of $\boldsymbol{\Omega}$ or $\boldsymbol{\Psi}$. Obtaining either of these from a known $\boldsymbol{\Omega}$ becomes increasingly difficult as N gets large.

For GLS to be feasible when N is very large, we cannot simply invert $\boldsymbol{\Omega}$ numerically and perhaps find $\boldsymbol{\Psi}$ numerically.

Sometimes $\boldsymbol{\Omega}$ depends on a short vector of known, or estimable, parameters, say γ. If so, we may be able to find an analytic expression for $\boldsymbol{\Psi}(\gamma)x$ for any vector x.

Then we can form $\boldsymbol{\Psi}^\top(\gamma)y$ and $\boldsymbol{\Psi}^\top(\gamma)X$ and compute the GLS estimator by running OLS on the transformed data:

$$\hat{\boldsymbol{\beta}}_{\text{GLS}} = \left(X^\top\boldsymbol{\Psi}(\gamma)\boldsymbol{\Psi}^\top(\gamma)X\right)^{-1}X^\top\boldsymbol{\Psi}(\gamma)\boldsymbol{\Psi}^\top(\gamma)y. \tag{16}$$

Alternatively, we can find an analytic expression for $x_1^\top\boldsymbol{\Omega}^{-1}(\gamma)x_2$ for any pair of vectors x_1 and x_2.

Then we can form $X^\top\boldsymbol{\Omega}^{-1}(\gamma)X$ and $X^\top\boldsymbol{\Omega}^{-1}(\gamma)y$ and compute

$$\hat{\boldsymbol{\beta}}_{\text{GLS}} = \left(X^\top\boldsymbol{\Omega}^{-1}(\gamma)X\right)^{-1}X^\top\boldsymbol{\Omega}^{-1}(\gamma)y. \tag{17}$$

Example. When the u_t follow an AR(1) process with autoregressive parameter ρ, the matrix $\boldsymbol{\Omega}(\rho)$ is easily seen to be

$$\mathbf{\Omega}(\rho) = \frac{\sigma^2}{1-\rho^2}\begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{N-1} \\ \rho & 1 & \rho & \cdots & \rho^{N-2} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho^{N-1} & \rho^{N-2} & \rho^{N-3} & \cdots & 1 \end{bmatrix}. \tag{18}$$

The corresponding $\mathbf{\Psi}(\rho)$ matrix is

$$\mathbf{\Psi}(\rho) = \begin{bmatrix} (1-\rho^2)^{1/2} & -\rho & 0 & \cdots & 0 & 0 \\ 0 & 1 & -\rho & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -\rho \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix}. \tag{19}$$

Notice that the transformation for the first observation differs from the transformation for all the other observations. See Beach and MacKinnon (Econometrica, 1978).

Whether ρ is known or estimated, the next step in GLS estimation of a linear regression model with AR(1) errors is to form the vector $\boldsymbol{\Psi}^\top \boldsymbol{y}$ and the matrix $\boldsymbol{\Psi}^\top \boldsymbol{X}$.

This can be done without having to store the $N \times N$ matrix $\boldsymbol{\Psi}$ in computer memory.

From (19), the first element of $\boldsymbol{\Psi}^\top \boldsymbol{y}$ is $(1 - \rho^2)^{1/2} y_1$, and the remaining elements are $y_t - \rho y_{t-1}$ for $t = 2, \ldots, N$.

Each column of $\boldsymbol{\Psi}^\top \boldsymbol{X}$ has precisely the same form as $\boldsymbol{\Psi}^\top \boldsymbol{y}$ and can be calculated in precisely the same way.

We could easily define a routine that takes a vector $\boldsymbol{x}$ and a scalar ρ and forms $\boldsymbol{\Psi}^\top \boldsymbol{x}$, then apply it to $\boldsymbol{y}$ and to each of the columns of $\boldsymbol{X}$.

When we either know ρ or can estimate it consistently, we just have to regress $\boldsymbol{\Psi}^\top \boldsymbol{y}$ on $\boldsymbol{\Psi}^\top \boldsymbol{X}$ in order to obtain (feasible) GLS estimates of $\boldsymbol{\beta}$.

For maximum likelihood estimation, we choose $\hat{\rho}$ to maximize the loglikelihood function.

Weighted Least Squares

Let $\mathbf{\Omega}$ be diagonal, and let ω_i^2 denote the i^{th} diagonal element of $\mathbf{\Omega}$.

Then the GLS regression is just

$$\omega_i^{-1} y_i = \omega_i^{-1} \boldsymbol{X}_i \boldsymbol{\beta} + \omega_i^{-1} u_i. \tag{20}$$

The regressand and regressors are ω_i^{-1} times the dependent and independent variables, and the variance of $\omega_i^{-1} u_i$ is 1.

This special case of GLS is called **weighted least squares**, or **WLS**.

Observations for which the ω_i^2 is large are given low weights, and observations for which it is small are given high weights.

If $\mathbf{\Omega} = \sigma^2 \mathbf{\Delta}$, with $\mathbf{\Delta}$ known but σ^2 unknown, (20) remains valid if we reinterpret ω_i^2 as the i^{th} diagonal element of $\mathbf{\Delta}$. The variance of the transformed disturbances is now σ^2 instead of 1.

If the original regression is

$$y_i = \beta_1 + \beta_2 x_i + u_i, \tag{21}$$

then the weighted regression is

$$y_i/\omega_i = \beta_1(1/\omega_i) + \beta_2(x_i/\omega_i) + u_i/\omega_i. \tag{22}$$

The regressand is y_i/ω_i, the regressor that corresponds to the constant term is $1/\omega_i$, and the regressor that corresponds to x_i is x_i/ω_i.

Transforming the data so that the transformed disturbances have the same variance is not the only reason for weighting.

Sometimes, **sample weights** are used when different observations represent different numbers of observations in the population.

Using sample weights induces heteroskedasticity instead of correcting for it, so we need to use an HCCME or CRVE.

Feasible Generalized Least Squares

In practice, $\mathbf{\Omega}$ is often not known even up to a scalar factor.

But in many cases $\mathbf{\Omega}$, or $\mathbf{\Delta}$, is assumed to depend in a known way on a vector of unknown parameters γ.

If we can estimate γ consistently, we can obtain $\mathbf{\Omega}(\hat{\gamma})$ and $\mathbf{\Psi}(\hat{\gamma})$.

Then we can compute GLS estimates conditional on $\mathbf{\Psi}(\hat{\gamma})$. This is called **feasible generalized least squares**, or **feasible GLS**.

As a simple example, consider the linear regression model

$$y_i = \boldsymbol{X}_i\boldsymbol{\beta} + u_i, \quad \mathrm{E}(u_i^2) = \exp(\boldsymbol{Z}_i\boldsymbol{\gamma}). \tag{23}$$

Here $\exp(\boldsymbol{Z}_i\boldsymbol{\gamma})$ is a **skedastic function**.

It is designed to be positive for every γ, so that we do not have to worry about negative estimated variances.

In order to estimate γ consistently, we must first obtain consistent estimates of the u_i.

We start by running OLS to obtain $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{u}}$.

We then run the auxiliary linear regression

$$\log \hat{u}_i^2 = \boldsymbol{Z}_i \boldsymbol{\gamma} + v_i, \quad i = 1, \ldots, N, \tag{24}$$

to find the OLS estimates $\hat{\gamma}$.

These estimates are then used to compute

$$\hat{\omega}_i = \left(\exp(\boldsymbol{Z}_i \hat{\boldsymbol{\gamma}})\right)^{1/2} \quad \text{for all } i. \tag{25}$$

Finally, feasible GLS estimates of $\boldsymbol{\beta}$ are obtained by using OLS to estimate the regression

$$y_i / \hat{\omega}_i = (1/\hat{\omega}_i) \boldsymbol{X}_i \boldsymbol{\beta} + \epsilon_i. \tag{26}$$

If we substitute $X\beta_0 + u$ for y into the formula for the GLS estimator, we find that

$$\hat{\beta}_{\text{GLS}} = \beta_0 + (X^\top \Omega^{-1} X)^{-1} X^\top \Omega^{-1} u. \tag{27}$$

Taking β_0 to the l.h.s., multiplying each factor by an appropriate power of N, and taking probability limits, we see that

$$N^{1/2}(\hat{\beta}_{\text{GLS}} - \beta_0) \stackrel{a}{=} \left(\operatorname*{plim}_{N\to\infty} \frac{1}{N} X^\top \Omega^{-1} X \right)^{-1} N^{-1/2} X^\top \Omega^{-1} u. \tag{28}$$

Under standard assumptions, the inverse matrix is a nonstochastic $k \times k$ matrix with full rank. The vector is stochastic and follows a multivariate normal distribution.

The asymptotic distribution of $N^{1/2}(\hat{\beta}_{\text{GLS}} - \beta_0)$ follows from (28). It is multivariate normal with covariance matrix

$$\text{Var}\left(N^{1/2}(\hat{\beta}_{\text{GLS}} - \beta_0)\right) = \left(\operatorname*{plim}_{N\to\infty} \frac{1}{N} X^\top \Omega^{-1} X \right)^{-1}. \tag{29}$$

Now consider the analog of (28) for feasible GLS,

$$N^{1/2}(\hat{\boldsymbol{\beta}}_{\mathrm{F}} - \boldsymbol{\beta}_0) \stackrel{a}{=} \left(\operatorname*{plim}_{N\to\infty} \frac{1}{N} \boldsymbol{X}^\top \boldsymbol{\Omega}^{-1}(\hat{\gamma}) \boldsymbol{X}\right)^{-1} N^{-1/2} \boldsymbol{X}^\top \boldsymbol{\Omega}^{-1}(\hat{\gamma}) \boldsymbol{u}. \tag{30}$$

It is not trivial to prove that $\hat{\boldsymbol{\beta}}_{\mathrm{F}}$ is asymptotically equivalent to $\hat{\boldsymbol{\beta}}_{\mathrm{GLS}}$, but it can be done if $\hat{\gamma}$ is root-N consistent.

How well FGLS performs depends on how well $\hat{\gamma}$ estimates γ and how sensitive $\hat{\boldsymbol{\beta}}_{\mathrm{GLS}}$ is to variation in $\boldsymbol{\Omega}(\gamma)$.

- When $\boldsymbol{\Omega}(\hat{\gamma})$ is a very good estimator, then feasible GLS has essentially the same properties as GLS itself.
- In addition, the FGLS covariance matrix estimator will be similar to the GLS one.
- But if $\boldsymbol{\Omega}(\hat{\gamma})$ estimates $\boldsymbol{\Omega}(\gamma)$ poorly, feasible GLS estimates may have quite different properties from real GLS estimates, and inferences based on them may be quite misleading.

Testing for Heteroskedasticity

Testing for heteroskedasticity is no longer very popular. Unless N is small, it makes sense to use an HCCME without bothering to test.

A reasonably general model of conditional heteroskedasticity is

$$\mathrm{E}(u_i^2 \mid \Omega_i) = h(\delta + \boldsymbol{Z}_i\boldsymbol{\gamma}), \tag{31}$$

where $h(\cdot)$ is a nonlinear skedastic function, $\boldsymbol{Z}_i$ is a $1 \times r$ vector of exogenous or predetermined variables belonging to the information set Ω_i, δ is a scalar parameter, and $\boldsymbol{\gamma}$ is an r-vector of parameters.

One plausible specification of the skedastic function is

$$h(\delta + \boldsymbol{Z}_i\boldsymbol{\gamma}) = \exp(\delta + \boldsymbol{Z}_i\boldsymbol{\gamma}) = \exp(\delta)\exp(\boldsymbol{Z}_i\boldsymbol{\gamma}). \tag{32}$$

This reduces to $\sigma^2 \equiv \exp(\delta)$ when $\boldsymbol{\gamma} = \mathbf{0}$.

If we define v_i as $u_i^2 - \mathrm{E}(u_i^2)$, then equation (31) implies that

$$u_i^2 = h(\delta + \boldsymbol{Z}_i\boldsymbol{\gamma}) + v_i. \tag{33}$$

This looks like a nonlinear regression model, although the v_i must be strongly skewed to the right.

We can test for $\boldsymbol{\gamma} = \mathbf{0}$ by running the linear regression

$$\hat{u}_i^2 = b_\delta + \boldsymbol{Z}_i\boldsymbol{b}_\gamma + \text{residual}. \tag{34}$$

This is an easy way to test for all sorts of heteroskedasticity. We just have to decide on $\boldsymbol{Z}_i$.

The test statistic is usually either $F \overset{a}{\sim} F(r, N-r-1)$ or $NR^2 \overset{a}{\sim} \chi^2(r)$.

The theoretical basis for (34) is the **Gauss-Newton regression**, to be discussed in the context of nonlinear least squares (NLS).

Finite-sample properties of tests based on (34) are not great, especially when r is not small, so it is good to use bootstrap P values.

Models for Panel Data

Panel data involve two dimensions, one of them time. Consider the linear regression model

$$y_{gt} = \boldsymbol{X}_{gt}\boldsymbol{\beta} + u_{gt}, \quad g = 1, \ldots, G, \;\; t = 1, \ldots, T, \tag{35}$$

where $\boldsymbol{X}_{gt}$ is a $1 \times k$ vector of observations on explanatory variables. There are G cross-sectional units (groups) and T time periods

In (35), there is just one observation per group-period pair. It is a **balanced panel** with $N = GT$ observations. (maybe i instead of g)

There can also be models with multiple observations, say N_{gt}, per group-period pair. Such an **unbalanced panel** will have $N = \sum_{g=1}^{G} N_g = \sum_{g=1}^{G} \sum_{t=1}^{T} N_{gt}$ observations. (i, g, and t subcripts)

Some panels are short and wide (e.g. $T = 5$, all N_g large), while others are tall and narrow (e.g. $T = 200$, $G = 2$). The former may be called **landscape** and the latter **portrait**.

OLS estimation of (35) is consistent but inefficient if the u_{gt} are correlated across g or across t.

The simplest and most popular way to allow for correlation across both g and t is the **error-components model**

$$u_{gt} = e_t + v_g + \epsilon_{gt}. \tag{36}$$

We can treat the e_t and v_g as either fixed or random. When we treat them as **fixed effects**, they are parameters to be estimated by OLS.

When we treat them as **random effects**, we need to use GLS.

The model (36) involves **two-way** fixed or random effects. For simplicity, we will focus on models with **one-way** effects.

We have already discussed linear regression models with fixed effects. For the model

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{D\eta} + \boldsymbol{u}, \tag{37}$$

with G cross-section fixed effects, we found that

$$\hat{\boldsymbol{\beta}}_{\text{FE}} = (\boldsymbol{X}^\top \boldsymbol{M}_{\boldsymbol{D}} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{M}_{\boldsymbol{D}} \boldsymbol{y}. \tag{38}$$

and that

$$\hat{\boldsymbol{\eta}} = \begin{bmatrix} \bar{y}_1 - \bar{\boldsymbol{X}}_1 \hat{\boldsymbol{\beta}}_{\text{FE}} \\ \bar{y}_2 - \bar{\boldsymbol{X}}_2 \hat{\boldsymbol{\beta}}_{\text{FE}} \\ \vdots \\ \bar{y}_G - \bar{\boldsymbol{X}}_G \hat{\boldsymbol{\beta}}_{\text{FE}} \end{bmatrix}. \tag{39}$$

Here the $\bar{y}_g$ and $\bar{\boldsymbol{X}}_g$ are sample means over each cross-section group.

Although we lose efficiency by using OLS, we can make asymptotically valid inferences by using a CRVE with (37).

We could (and often should) also add time fixed effects.

The fixed-effects estimator (38) is also called the **within-groups estimator** because it depends entirely on within-group variation.

The FE estimator is inefficient. It ignores between-group variation.

When any explanatory variables in $\boldsymbol{X}$ are well explained by $\boldsymbol{D}$, the corresponding elements of $\hat{\boldsymbol{\beta}}_{\text{FE}}$ have large variances.

Because we condition on $\boldsymbol{D}$, it is OK for the fixed effects to be correlated with $\boldsymbol{X}$.

The FE estimator cannot be used with any explanatory variable that varies only across groups. Such a variable would be collinear with $\boldsymbol{D}$.

It is often impossible to include GT group-time effects.

For random-effects estimation, we must assume that

$$\mathrm{E}(u_{gt} \,|\, \boldsymbol{X}) = \mathrm{E}(v_g + \epsilon_{gt} \,|\, \boldsymbol{X}) = 0, \tag{40}$$

since v_g and ϵ_{gt} are then both independent of $\boldsymbol{X}$. Strong assumption!

In a panel of observations on workers, many variables might be correlated with an unobserved variable like ability, which implicitly enters into the individual-specific disturbance v_g.

To estimate a random-effects model, we use (feasible) GLS.

$$\text{Var}(u_{gt}) = \sigma_v^2 + \sigma_\epsilon^2, \tag{41}$$

$$\text{Cov}(u_{gt}\, u_{gs}) = \sigma_v^2, \text{ and} \tag{42}$$

$$\text{Cov}(u_{gt}\, u_{hs}) = 0 \text{ for all } g \neq h. \tag{43}$$

These define the elements of the $N \times N$ covariance matrix $\mathbf{\Omega}$.

If the data are ordered by the cross-sectional units in G blocks of T observations each, this matrix has the form

$$\mathbf{\Omega} = \begin{bmatrix} \mathbf{\Sigma} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{\Sigma} \end{bmatrix}, \quad \mathbf{\Sigma} \equiv \sigma_\epsilon^2 \mathbf{I}_T + \sigma_v^2 \boldsymbol{\iota}\boldsymbol{\iota}^\top, \tag{44}$$

where $\mathbf{\Sigma}$ is the $T \times T$ matrix with $\sigma_v^2 + \sigma_\epsilon^2$ in every position on the principal diagonal and σ_v^2 everywhere else.

To obtain $\hat{\boldsymbol{\beta}}_{\mathrm{RE}}$, we need to estimate σ_ϵ^2 and σ_v^2 consistently; see ETM.

There is a tricky way to compute $\hat{\boldsymbol{\beta}}_{\mathrm{RE}}$. First, define $\boldsymbol{P_D} \equiv \mathbf{I} - \boldsymbol{M_D}$.

Then compute

$$\hat{\lambda} = 1 - \left(\frac{T\hat{\sigma}_v^2}{\hat{\sigma}_\epsilon^2} + 1 \right)^{-1/2}, \tag{45}$$

and run the OLS regression

$$(\mathbf{I} - \hat{\lambda}\boldsymbol{P_D})\boldsymbol{y} = (\mathbf{I} - \hat{\lambda}\boldsymbol{P_D})\boldsymbol{X}\boldsymbol{\beta} + \text{residuals}. \tag{46}$$

This is not hard, because we do not actually need $\boldsymbol{P_D}$. The vector $\boldsymbol{P_D x}$ simply contains the G means of $\boldsymbol{x}$, spread out over all N observations.

The RE estimator from (46) is just a matrix-weighted average of the OLS estimator $\hat{\boldsymbol{\beta}}$ and the **between-groups** estimator

$$\hat{\boldsymbol{\beta}}_{\mathrm{BG}} = (\boldsymbol{X}^\top \boldsymbol{P_D} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{P_D} \boldsymbol{y}, \tag{47}$$

which uses only G observations.