Generalized Regression Model

Based on Greene's Note 15 (Chapter 8)

Generalized Regression Model

Setting: The classical linear model assumes that $E[\boldsymbol{\varepsilon\varepsilon}'] = Var[\boldsymbol{\varepsilon}] = \sigma^2\mathbf{I}$. That is, observations are uncorrelated and all are drawn from a distribution with the same variance. The **generalized regression** (**GR**) model allows the variances to differ across observations and allows correlation across observations.

Implications

- The assumption that Var[$\boldsymbol{\varepsilon}$] = $\sigma^2\mathbf{I}$ is used to derive the result Var[**b**] = $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$. If it is not true, then the use of $s^2(\mathbf{X}'\mathbf{X})^{-1}$ to estimate Var[**b**] is inappropriate.
- The assumption was used to derive most of our test statistics, so they must be revised as well.
- Least squares gives each observation a weight of 1/n. But, if the variances are not equal, then some observations are more informative than others.
- Least squares is based on simple sums, so the information that one observation might provide about another is never used.

GR Model

- The generalized regression model:

  $\mathbf{y} = \mathbf{X}\beta + \boldsymbol{\varepsilon}$,

  $E[\boldsymbol{\varepsilon}|\mathbf{X}] = \mathbf{0}$, $Var[\boldsymbol{\varepsilon}|\mathbf{X}] = \sigma^2\boldsymbol{\Omega}$.

  Regressors are well behaved. We consider some examples with Trace $\boldsymbol{\Omega}$ = n. (This is a normalization with no content.)

- Leading Cases
  - Simple heteroscedasticity
  - Autocorrelation
  - Panel data and heterogeneity more generally.

Least Squares

- Still **unbiased**. (Proof did not rely on Ω)
- For **consistency**, we need the true variance of **b**,

$$\begin{aligned}\text{Var}[\mathbf{b}|\mathbf{X}] &= E[(\mathbf{b}-\boldsymbol{\beta})(\mathbf{b}-\boldsymbol{\beta})'|\mathbf{X}] \\ &= (\mathbf{X}'\mathbf{X})^{-1} E[\mathbf{X}'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\mathbf{X}] (\mathbf{X}'\mathbf{X})^{-1} \\ &= \boldsymbol{\sigma}^2 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\boldsymbol{\Omega}\mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} .\end{aligned}$$

Divide all 4 terms by *n*. If the middle one converges to a finite matrix of constants, we have the result, so we need to examine

$$(1/n)\mathbf{X}'\boldsymbol{\Omega}\mathbf{X} = (1/n)\Sigma_i\Sigma_j \ \omega_{ij} \mathbf{x}_i \mathbf{x}_j'.$$

This will be another assumption of the model.

- **Asymptotic normality**? Easy for heteroscedasticity case, very difficult for autocorrelation case.

Robust Covariance Matrix

- Robust estimation:
- How to estimate $\text{Var}[\mathbf{b}|\mathbf{X}] = \sigma^2 (\mathbf{X'X})^{-1} \mathbf{X'\Omega X} (\mathbf{X'X})^{-1}$ for the LS $\mathbf{b}$?
- The distinction between estimating

  $\sigma^2\boldsymbol{\Omega}$ an n by n matrix

  and estimating

  $\sigma^2 \mathbf{X'\Omega X} = \sigma^2 \Sigma_i \Sigma_j \omega_{ij}\, \mathbf{x}_i\, \mathbf{x}_j'$
- For modern applied econometrics,
  - The White estimator
  - Newey-West.

The White Estimator

$$\text{Est.Var}[\mathbf{b}] = (\mathbf{X'X})^{-1}\left[\sum_{i=1}^{n} e_i^2 \mathbf{x}_i \mathbf{x}_i'\right](\mathbf{X'X})^{-1}$$

Use $$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} e_i^2}{n}$$

$$\hat{\omega}_i = \frac{n e_i^2}{\hat{\sigma}^2},\ \hat{\mathbf{\Omega}} = \text{diag}(\omega_i) \text{ note } \text{tr}(\mathbf{\Omega}) = n$$

$$\text{Est.Var}[\mathbf{b}] = \frac{\hat{\sigma}^2}{n}\left(\frac{\mathbf{X'X}}{n}\right)^{-1}\left(\frac{\mathbf{X'\hat{\Omega}X}}{n}\right)\left(\frac{\mathbf{X'X}}{n}\right)$$

Does $$\hat{\sigma}^2\left(\frac{\mathbf{X'\hat{\Omega}X}}{n}\right) - \sigma^2\left(\frac{\mathbf{X'\Omega X}}{n}\right) \to \mathbf{0?}$$

Newey-West Estimator

Heteroscedasticity Component - Diagonal Elements

$$\mathbf{S}_0 = \frac{1}{n}\sum_{i=1}^{n} e_i^2 \mathbf{x}_i \mathbf{x}_i'$$

Autocorrelation Component - Off Diagonal Elements

$$\mathbf{S}_1 = \frac{1}{n}\sum_{l=1}^{L}\sum_{t=l+1}^{n} w_l e_t e_{t-l}(\mathbf{x}_t \mathbf{x}_{t-l}' + \mathbf{x}_{t-l}\mathbf{x}_t')$$

$$w_l = 1 - \frac{l}{L+1} = \text{"Bartlett weight"}$$

$$\text{Est.Var}[\mathbf{b}] = \frac{1}{n}\left(\frac{\mathbf{X'X}}{n}\right)^{-1}[\mathbf{S}_0 + \mathbf{S}_1]\left(\frac{\mathbf{X'X}}{n}\right)^{-1}$$

Generalized Least Squares

A transformation of the model:

$$\mathbf{P} = \Omega^{-1/2}.\ \mathbf{P'P} = \Omega^{-1}$$

$$\mathbf{Py} = \mathbf{PX}\beta + \mathbf{P}\varepsilon \quad \text{or}$$

$$\mathbf{y}^* = \mathbf{X}^*\beta + \varepsilon^*. \quad \text{Why?}$$

$$\mathrm{E}[\varepsilon^*\varepsilon^{*\prime}|\mathbf{X}^*] = \mathbf{P}\mathrm{E}[\varepsilon\varepsilon'|\mathbf{X}^*]\mathbf{P'}$$

$$= \mathbf{P}\mathrm{E}[\varepsilon\varepsilon'|\mathbf{X}]\mathbf{P'}$$

$$= \sigma^2\mathbf{P}\Omega\mathbf{P'} = \sigma^2\,\Omega^{-1/2}\,\Omega\Omega^{-1/2} = \sigma^2\,\Omega^0$$

$$= \sigma^2\mathbf{I}$$

Generalized Least Squares

Aitken theorem. The **Generalized Least Squares** estimator, GLS.

$$\mathbf{Py} = \mathbf{PX}\beta + \mathbf{P}\varepsilon \quad \text{or}$$

$$\mathbf{y}^* = \mathbf{X}^*\beta + \varepsilon^*. \quad E[\varepsilon^*\varepsilon^{*\prime}|\mathbf{X}^*] = \sigma^2\mathbf{I}$$

Use ordinary least squares in the transformed model. Satisfies the Gauss – Markov theorem.

$$\mathbf{b}^* = (\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\mathbf{y}^*$$

Generalized Least Squares

Efficient estimation of β and, by implication, the inefficiency of least squares **b**.

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

$$= (\mathbf{X}'\mathbf{P}'\mathbf{P}\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{P}'\mathbf{P}\mathbf{y}$$

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

$\hat{\boldsymbol{\beta}} \neq \mathbf{b}$. $\hat{\boldsymbol{\beta}}$ is efficient,

so by construction, **b** is not.

Asymptotics for GLS

Asymptotic distribution of GLS. (NOTE: We apply the full set of results of the classical model to the transformed model).

- Unbiasedness
- Consistency - “well behaved data”
- Asymptotic distribution
- Test statistics

Unbiasedness

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

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

$$\mathrm{E}[\hat{\boldsymbol{\beta}} \mid \mathbf{X}] = \boldsymbol{\beta} + (\mathbf{X'\Omega^{-1}X})^{-1}\mathbf{X'\Omega^{-1}}\mathrm{E}[\boldsymbol{\varepsilon} \mid \mathbf{X}]$$

$$= \boldsymbol{\beta} \quad \text{if} \quad \mathrm{E}[\boldsymbol{\varepsilon} \mid \mathbf{X}] = \mathbf{0}$$

Consistency

Use Mean Square

$$\text{Var}[\hat{\boldsymbol{\beta}}|\mathbf{X}]=\frac{\sigma^2}{n}\left(\frac{\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}}{n}\right)^{-1}\rightarrow \mathbf{0}\,?$$

Requires to be $\left(\frac{\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}}{n}\right)$ "well behaved"

Either converge to a constant matrix or diverge.

Heteroscedasticity case:

$$\frac{\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}}{n}=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\omega_{ii}}\mathbf{x}_i\mathbf{x}_i'$$

Autocorrelation case:

$$\frac{\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}}{n}=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{\omega_{ij}}\mathbf{x}_i\mathbf{x}_j'.$$ n^2 terms. Convergence is unclear.

Asymptotic Normality

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \sqrt{n}\left(\frac{\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}}{n}\right)^{-1}\frac{1}{n}\mathbf{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon}$$

Converge to normal with a stable variance O(1)?

$$\left(\frac{\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}}{n}\right)^{-1} \rightarrow \text{ a constant matrix?}$$

$$\frac{1}{n}\mathbf{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon} \rightarrow \text{ a mean to which we can apply the central limit theorem?}$$

Heteroscedasticity case?

$$\frac{1}{n}\mathbf{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon} = \frac{1}{n}\sum_{i=1}^{n}\frac{\mathbf{x}_i}{\sqrt{\omega_i}}\left(\frac{\varepsilon_i}{\sqrt{\omega_i}}\right). \text{ Var}\left(\frac{\varepsilon_i}{\sqrt{\omega_i}}\right) = \sigma^2, \frac{\mathbf{x}_i}{\sqrt{\omega_i}} \text{ is just data.}$$

Apply Lindeberg-Feller. (Or assuming $\mathbf{x}_i / \sqrt{\omega_i}$ is a draw from a common distribution with mean and fixed variance - some recent treatments.)

Autocorrelation case?

Asymptotic Normality (Cont.)

For the autocorrelation case

$$\frac{1}{n}\mathbf{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\boldsymbol{\Omega}^{\mathbf{ij}}\,\mathbf{x}_i\mathbf{x}_j\varepsilon_i\varepsilon_i$$

Does the double sum converge? Uncertain. Requires elements of $\boldsymbol{\Omega}^{-1}$ to become small as the distance between i and j increases. (Has to resemble the heteroscedasticity case.)

Test Statistics
(Assuming Known Ω)

- With known Ω, apply all familiar results to the transformed model.
- With normality, t and F statistics apply to least squares based on **Py** and **PX**
- With asymptotic normality, use Wald statistics and the chi-squared distribution, still based on the transformed model.

Generalized (Weighted) Least Squares:Heteroscedasticity

$$\text{Var}[\varepsilon] = \sigma^2\Omega = \sigma^2 \begin{bmatrix} \omega_1 & 0 & \dots & 0 \\ 0 & \omega_2 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & \omega_n \end{bmatrix}$$

$$\Omega^{-1/2} = \begin{bmatrix} 1/\sqrt{\omega_1} & 0 & \dots & 0 \\ 0 & 1/\sqrt{\omega_2} & \dots & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 1/\sqrt{\omega_n} \end{bmatrix}$$

$$\hat{\boldsymbol{\beta}} = (\mathbf{X'\Omega^{-1}X})^{-1}(\mathbf{X'\Omega^{-1}y}) = \left(\sum_{i=1}^{n} \frac{1}{\omega_i}\mathbf{x_i x}_i'\right)^{-1}\left(\sum_{i=1}^{n} \frac{1}{\omega_i}\mathbf{x_i}y_i\right)$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}\left(\frac{y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}}}{\omega_i}\right)^2}{n-K}$$

Autocorrelation

$\varepsilon_t = \rho\varepsilon_{t-1} + u_t$

('First order autocorrelation.' How does this come about?)

Assume $-1 < \rho < 1$. Why?

u_t = 'nonautocorrelated white noise'

$\varepsilon_t = \rho\varepsilon_{t-1} + u_t$ (the autoregressive form)

$= \rho(\rho\varepsilon_{t-2} + u_{t-1}) + u_t$

= ... (continue to substitute)

$= u_t + \rho u_{t-1} + \rho^2 u_{t-2} + \rho^3 u_{t-3} + \ldots$

= (the moving average form)

(Some observations about modeling time series.)

Autocorrelation

$$\mathrm{Var}[\varepsilon_t] = \mathrm{Var}[u_t + \rho u_{t-1} + \rho^2 u_{t-1} + \ldots]$$

$$= \mathrm{Var}\left[\sum_{i=0}^{\infty} \rho^i u_{t-i}\right]$$

$$= \sum_{i=0}^{\infty} \rho^{2i} \sigma_u^2 = \frac{\sigma_u^2}{1-\rho^2}$$

An easier way: Since $\mathrm{Var}[\varepsilon_t] = \mathrm{Var}[\varepsilon_{t-1}]$ and $\varepsilon_t = \rho\varepsilon_{t-1} + u_t$

$$\mathrm{Var}[\varepsilon_t] = \rho^2 \mathrm{Var}[\varepsilon_{t-1}] + \mathrm{Var}[u_t] + 2\rho \mathrm{Cov}[\varepsilon_{t-1}, u_t]$$

$$= \rho^2 \mathrm{Var}[\varepsilon_t] + \sigma_u^2$$

$$= \frac{\sigma_u^2}{1-\rho^2}$$

Autocovariances

Continuing...

$$\begin{aligned}
\text{Cov}[\varepsilon_t, \varepsilon_{t-1}] &= \text{Cov}[\rho\varepsilon_{t-1} + u_t, \varepsilon_{t-1}] \\
&= \rho\text{Cov}[\varepsilon_{t-1}, \varepsilon_{t-1}] + \text{Cov}[u_t, \varepsilon_{t-1}] \\
&= \rho\text{Var}[\varepsilon_{t-1}] = \rho\text{Var}[\varepsilon_t] \\
&= \frac{\rho\sigma_u^2}{(1-\rho^2)}
\end{aligned}$$

$$\begin{aligned}
\text{Cov}[\varepsilon_t, \varepsilon_{t-2}] &= \text{Cov}[\rho\varepsilon_{t-1} + u_t, \varepsilon_{t-2}] \\
&= \rho\text{Cov}[\varepsilon_{t-1}, \varepsilon_{t-2}] + \text{Cov}[u_t, \varepsilon_{t-2}] \\
&= \rho\text{Cov}[\varepsilon_t, \varepsilon_{t-1}] \\
&= \frac{\rho^2\sigma_u^2}{(1-\rho^2)} \text{ and so on.}
\end{aligned}$$

Autocorrelation Matrix

$$\sigma^2\boldsymbol{\Omega} = \left(\frac{\sigma_u^2}{1-\rho^2}\right)\begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{T-1} \\ \rho & 1 & \rho & \cdots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \cdots & 1 \end{bmatrix}$$

(Note, trace $\boldsymbol{\Omega}$ = n as required.)

Generalized Least Squares

$$\mathbf{\Omega}^{-1/2} = \begin{bmatrix} \sqrt{1-\rho^2} & 0 & 0 & \dots & 0 \\ -\rho & 1 & 0 & \dots & 0 \\ 0 & -\rho & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & -\rho & 0 \end{bmatrix}$$

$$\mathbf{\Omega}^{-1/2}\mathbf{y} = \begin{pmatrix} \left(\sqrt{1-\rho^2}\right) y_1 \\ y_2 - \rho y_2 \\ y_3 - \rho y_2 \\ \dots \\ y_T - \rho_{T-1} \end{pmatrix}$$

The Autoregressive Transformation

$$y_t = \mathbf{x_t}'\boldsymbol{\beta} + \varepsilon_t \qquad \varepsilon_t = \rho\varepsilon_{t-1} + u_t$$

$$\rho y_{t-1} = \rho\mathbf{x_{t-1}}'\boldsymbol{\beta} + \rho\varepsilon_{t-1}$$

– – – – – – – – – – – – –

$$y_t - \rho y_{t-1} = (\mathbf{x_t} - \rho\mathbf{x_{t-1}})'\boldsymbol{\beta} + (\varepsilon_t - \rho\varepsilon_{t-1})$$

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

(Where did the first observation go?)

Unknown Ω

- The problem (of course), Ω is unknown. For now, we will consider two methods of estimation:
  - Two step, or feasible estimation. Estimate Ω first, then do GLS. Emphasize - same logic as White and Newey-West. We don't need to estimate Ω. We need to find a matrix that behaves the same as $(1/n)\mathbf{X}'\Omega^{-1}\mathbf{X}$.
  - Properties of the feasible GLS estimator
- Maximum likelihood estimation of β, σ^2, and Ω all at the same time.
  - Joint estimation of all parameters. Fairly rare. Some generalities…
  - We will examine two applications: Harvey's model of heteroscedasticity and Beach-MacKinnon on the first order autocorrelation model

Specification

- Ω must be specified first.
- A full unrestricted Ω contains n(n+1)/2 - 1 parameters. (Why minus 1? Remember, tr(Ω) = n, so one element is determined.)
- Ω is generally specified in terms of a few parameters. Thus, $\Omega = \Omega(\theta)$ for some small parameter vector θ. It becomes a question of estimating θ.
- Examples:

Heteroscedasticity: Harvey's Model

- $\text{Var}[\varepsilon_i \mid \mathbf{X}] = \sigma^2 \exp(\boldsymbol{\gamma}'\mathbf{z}_i)$
- $\text{Cov}[\varepsilon_i,\varepsilon_j \mid \mathbf{X}] = 0$

  e.g.: z_i = firm size

  e.g.: $\mathbf{z}_i$ = a set of dummy variables (e.g., countries) (The groupwise heteroscedasticity model.)
- $[\sigma^2 \boldsymbol{\Omega}]$ = diagonal $[\exp(\theta + \boldsymbol{\gamma}'\mathbf{z}_i)]$,

  $\theta = \log(\sigma^2)$

AR(1) Model of Autocorrelation

$$\sigma^2\boldsymbol{\Omega} = \left(\frac{\sigma_u^2}{1-\rho^2}\right)\begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{T-1} \\ \rho & 1 & \rho & \cdots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \cdots & 1 \end{bmatrix}$$

Two Step Estimation

The general result for estimation when $\boldsymbol{\Omega}$ is estimated.

GLS uses $[\mathbf{X}'\Omega^{-1}\mathbf{X}]\mathbf{X}'\,\Omega^{-1}\,\mathbf{y}$ which converges in probability to $\boldsymbol{\beta}$.

We seek a vector which converges to the same thing that this does. Call it “Feasible GLS” or FGLS, based on $[\mathbf{X}'\hat{\boldsymbol{\Omega}}^{-1}\mathbf{X}]\mathbf{X}'\,\hat{\boldsymbol{\Omega}}^{-1}\mathbf{y}$

The object is to find a set of parameters such that

$$[\mathbf{X}'\hat{\boldsymbol{\Omega}}^{-1}\mathbf{X}]\mathbf{X}'\hat{\boldsymbol{\Omega}}^{-1}\mathbf{y} \;-\; [\mathbf{X}'\,\Omega^{-1}\mathbf{X}]\mathbf{X}'\,\Omega^{-1}\mathbf{y} \;\rightarrow\; \mathbf{0}$$

Feasible GLS

For FGLS estimation, we do not seek an estimator of $\mathbf{\Omega}$ such that

$$\hat{\mathbf{\Omega}} - \mathbf{\Omega} \to \mathbf{0}$$

This makes no sense, since $\hat{\mathbf{\Omega}}$ is nxn and does not "converge" to anything. We seek a matrix $\mathbf{\Omega}$ such that

$$(1/n)\mathbf{X'}\hat{\mathbf{\Omega}}^{-1}\mathbf{X} - (1/n)\mathbf{X'}\mathbf{\Omega}^{-1}\mathbf{X} \to \mathbf{0}$$

For the asymptotic properties, we will require that

$$(1/n)\mathbf{X'}\hat{\mathbf{\Omega}}^{-1}\varepsilon - (1/n)\mathbf{X'}\mathbf{\Omega}^{-1}\varepsilon \to \mathbf{0}$$

Note in this case, these are two random vectors, which we require to converge to the same random vector.

Two Step FGLS

(Theorem 8.5) To achieve full efficiency, we do not need an **efficient** estimate of the parameters in $\mathbf{\Omega}$, only a consistent one. Why?

Harvey's Model

Examine Harvey's model once again.

Methods of estimation:

Two step FGLS: Use the least squares residuals to estimate θ, then use

$\{\mathbf{X}'[\boldsymbol{\Omega}(\theta)]^{-1}\mathbf{X}\}^{-1}\mathbf{X}'[\boldsymbol{\Omega}(\theta)]^{-1}\mathbf{y}$ to estimate β.

Full maximum likelihood estimation. Estimate all parameters simultaneously.

A handy result due to Oberhofer and Kmenta - the "zig-zag" approach.

Examine a model of groupwise heteroscedasticity.

Harvey's Model for Groupwise Heteroscedasticity

Groupwise sample, y_{ig}, x_{ig},…

N groups, each with N_g observations.

$\text{Var}[\varepsilon_{ig}] = \sigma_g^2$

Let $d_{ig} = 1$ if observation i,g is in group j, 0 else.

= group dummy variable.

$\text{Var}[\varepsilon_{ig}] = \sigma_g^2 \exp(\theta_2 d_2 + \ldots \theta_G d_G)$

$\text{Var}_1 = \sigma_g^2$, $\text{Var}_2 = \sigma_g^2 \exp(\theta_2)$ and so on.

Estimating Variance Components

- OLS is still consistent:
- Est.$Var_1 = e_1'e_1/N_1$ estimates σ_g^2
- Est.$Var_2 = e_2'e_2/N_2$ estimates $\sigma_g^2 \exp(\theta_2)$
- Estimator of θ_2 is $\ln[(e_2'e_2/N_2)/(e_1'e_1/N_1)]$
- (1) Now use FGLS – weighted least squares
- Recompute residuals using WLS slopes
- (2) Recompute variance estimators
- Iterate to a solution… between (1) and (2)