Lecture 7-c
GLS & FGLS

Brooks (4^{th} edition): Chapter 5

 1

Review – Generalized Regression Model

- Now, we go back to the CLM Assumptions:

(**A1**) DGP: $\boldsymbol{y} = \mathbf{X}\,\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ is correctly specified.

(**A2**) E[$\boldsymbol{\varepsilon}$|$\mathbf{X}$] = **0**

(**A3'**) Var[$\boldsymbol{\varepsilon}$|$\mathbf{X}$] = $\boldsymbol{\Sigma}$ (sometimes written Var[$\boldsymbol{\varepsilon}$|$\mathbf{X}$] = $\sigma^2\boldsymbol{\Omega}$)

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1T} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2T} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{T1} & \sigma_{T2} & \cdots & \sigma_T^2 \end{bmatrix}$$ -a (TxT) symmetric matrix

(**A4**) **X** has full column rank – rank(**X**) = k –, where T ≥ k.

- This is the generalized regression model (GRM).

- OLS is still unbiased (& consistent). Can we still use OLS?

Review – GRM: True Variance for b

- Now, we have (**A3'**) Var[**ε** | **X**] = **Σ**
- The true variance of **b** under (**A3'**) should be:

$$\text{Var}_T[\mathbf{b}\,|\,\mathbf{X}] = (\mathbf{X'X})^{-1}\,\mathbf{X'\Sigma X}\,(\mathbf{X'X})^{-1}$$

Example: We compute the true variance for the simplest case, a regression with only one explanatory variable and heteroscedastic **ε**:

$$\boldsymbol{y} = \mathbf{X}\,\beta + \boldsymbol{\varepsilon}, \qquad \varepsilon_i \sim D(0, \sigma_i^2)$$

$$\Rightarrow \qquad \text{Var}_T[\mathbf{b}\,|\,\mathbf{X}] = \left(\frac{1}{\sum_i^T (x_i - \bar{x})^2}\right)^2 \sum_{i=1}^T \sigma_i^2 (x_i - \bar{x})^2.$$

If we compute the OLS variance, we see how both estimators differ:

$$\text{Var}[\mathbf{b}\,|\,\mathbf{X}] = \frac{\sigma^2}{\sum_i^T (x_i - \bar{x})^2} \neq \text{Var}_T[\mathbf{b}\,|\,\mathbf{X}].$$

Review – GRM: True Variance for b

- Under (**A3'**), the OLS estimator of Var[**b**|**X**] = s^2 $(\mathbf{X'X})^{-1}$ is *biased.*
- If we want to use OLS for inferences (say, with *t-test* or *F-test*), we need to estimate Var_T[**b**|**X**].
- That is, we need to estimate the unknown **Σ**. But, it has $T*(T+1)/2$ parameters. Too many parameters to estimate with T observations!
- We will not be estimating **Σ**. Impossible with T data points.
- We will estimate $\mathbf{X'\Sigma X} = \sum_{i=1}^T \sum_{j=1}^T \sigma_{ij}\, \boldsymbol{x}_i\, \boldsymbol{x}_j{}'$, a ($kxk$) matrix. That is, we are estimating $[k*(k+1)]/2$ elements.

Review – GRM: Robust Covariance Matrix

- This distinction is very important in modern applied econometrics:
  - – The White estimator
  - – The Newey-West estimator

- Both estimators produce a *consistent* estimator of $\text{Var}_T[\mathbf{b}|\mathbf{X}]$:

$$\text{Var}_T[\mathbf{b}|\mathbf{X}] = (\mathbf{X'X})^{-1}\, \mathbf{X'\Sigma X}\, (\mathbf{X'X})^{-1}$$

Since **b** consistently estimates **β**, the OLS residuals, ***e***, are also consistent estimators of **ε**. We use **e** to consistently estimate $\mathbf{X'\Sigma X}$.

In practice, we use w_i $(= x_i\, e_i)$ to estimate $\mathbf{X'\Sigma X}$.

Review – GRM: The Newey-West Estimator

- The White estimator simplifies the estimation since it only assumes heteroscedasticity. Then, **Σ** is a diagonal matrix, with elements σ_i^2.

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sigma_T^2 \end{bmatrix} \qquad \text{-a } (T\text{x}T) \text{ matrix}$$

Thus, we need to estimate: $\mathbf{Q}^* = (1/T)\, \mathbf{X'\Sigma X}$ -a (kxk) matrix

- We use e_i^2 to estimate σ_i^2. That is,

we estimate $\mathbf{Q}^* = (1/T) \sum_{i=1}^{T} \sigma_i^2\, \boldsymbol{x}_i\, \boldsymbol{x}_i'$

with $\mathbf{S}_0 = (1/T) \sum_{i=1}^{T} e_i^2\, \boldsymbol{x}_i\, \boldsymbol{x}_i'$

Review – GRM: The Newey-West Estimator

- Newey-West allow for both heteroscedasticy and autocorrelation.

(**A3'**) Var[$\boldsymbol{\varepsilon}$|**X**] = $\boldsymbol{\Sigma}$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1T} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2T} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{T1} & \sigma_{T2} & \cdots & \sigma_T^2 \end{bmatrix} \qquad \text{-a } (T\text{x}T) \text{ matrix}$$

Now, we need to estimate

$$\mathbf{Q}^* = (1/T)\, \mathbf{X}'\boldsymbol{\Sigma}\mathbf{X} = (1/T) \sum_{i=1}^{T}\sum_{j=1}^{T} \sigma_{ij}\, \boldsymbol{x}_i\, \boldsymbol{x}_j'$$
$$= (1/T) \sum_{i=1}^{T}\{\sigma_{i1}\, \boldsymbol{x}_i\, \boldsymbol{x}_1 + \sigma_{i2}\, \boldsymbol{x}_i\, \boldsymbol{x}_2 + \cdots + \sigma_{iT}\, \boldsymbol{x}_i\, \boldsymbol{x}_T\}$$

- Newey and West (1987) follow White (1980) to produce a HAC (*Heteroscedasticity and Autocorrelation Consistent*) estimator of $\mathbf{Q}^*$:

$$\mathrm{S}_T = (1/T) \sum_{i=1}^{T}\sum_{j=1}^{T} e_i e_j\, \boldsymbol{x}_i\, \boldsymbol{x}_j'$$

Review – GRM: The Newey-West Estimator

- Newey and West (1987) estimator of $\mathbf{Q}^*$:

$$\mathrm{S}_T = (1/T) \sum_{i=1}^{T}\sum_{j=1}^{T} e_i e_j\, \boldsymbol{x}_i\, \boldsymbol{x}_j'$$

Then,

Est. Var[**b**] = $(1/T)\, (\mathbf{X'X}/T)^{-1}\, \mathrm{S}_T\, (\mathbf{X'X}/T)^{-1}$

Example: Back to the simplest case, a regression with only one explanatory variable, but now with a heteroscedastic and autocorrelated error term. We estimate the "true" variance of **b** with:

$$\mathrm{Var}_T[\mathbf{b}|\mathbf{X}] = \left(\frac{1}{\sum_i^T (x_i - \bar{x})^2}\right)^2 \{\sum_{i=1}^{T} e_i^2 (x_i - \bar{x})^2 +$$
$$+ \sum_{i=1}^{T}\sum_{j=i+1}^{T} (x_i - \bar{x})\, e_i\, e_j (x_j - \bar{x})\}$$

We add the sum of the autocovariances of w_i (= $x_i\, e_i$) to the White estimator of $\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X}$. If (auto-) covar($w_i$, w_j) are positive, the NW estimator will be bigger than the White one. A common case.

Review – GRM: The Newey-West Estimator

• Two components for the NW HAC estimator:
(1) Start with Heteroscedasticity Component:

$\mathbf{S}_0 = (1/T)\sum_{i=1}^{T} e_i^2 \boldsymbol{x}_i \boldsymbol{x}_i'$ – the White estimator.

(2) Add the Autocorrelation Component, cutting sum short with L.

$\mathbf{S}_T = \mathbf{S}_0 + \frac{1}{T}\sum_{l=1}^{L} k(l)\sum_{t=l+1}^{T}(\boldsymbol{x}_{t-l}e_{t-l}e_t\boldsymbol{x}_t' + \boldsymbol{x}_t e_t e_{t-l}\boldsymbol{x}_{t-l}')$

where

$k(\frac{j}{L(T)}) = \frac{L+1-|j|}{L+1}$ –decaying weights (*Bartlett kernel*)

L is the cut-off lag, which is a function of T. (More data, longer L).

The weights are linearly decaying, suppose $L = 12$. Then,
$k(1) = 12/13 = 0.92308$
$k(2) = 11/13 = 0.84615$
$k(3) = 10/13 = 0.76923$

$k(\frac{j}{L(T)}) = \frac{13-|j|}{13}$

Review – GRM: The Newey-West Estimator

• $\mathbf{S}_T = \mathbf{S}_0 + (1/T)\sum_{l=1}^{L} k(l)\sum_{t=l+1}^{T}(\boldsymbol{x}_{t-l}e_{t-l}e_t\boldsymbol{x}_t' + \boldsymbol{x}_t e_t e_{t-l}\boldsymbol{x}_{t-l}')$

Then,

Est. $\text{Var}_T[\mathbf{b}|\mathbf{X}] = (1/T)(\mathbf{X'X}/T)^{-1}\,\mathbf{S}_T\,(\mathbf{X'X}/T)^{-1}$ –NW's HAC Var.

• Asymptotic inferences can be based on OLS $\mathbf{b}$ & Est. $\text{Var}_T[\mathbf{b}|\mathbf{X}]$. We can use the usual tests and distributions.

Example: Back to the simplest case, a regression with only one explanatory variable, but with a heteroscedastic and autocorrelated error term. Suppose we set L = **12**, then:

$$\text{Var}_T[\mathbf{b}|\mathbf{X}] = \left(\frac{1}{\sum_i^T (x_i-\bar{x})^2}\right)^2 \{\sum_{t=1}^{T} e_t^2 (x_t-\bar{x})^2 + \\ + \sum_{l=1}^{L=12}\left\{\frac{13-|j|}{13}\right\}\sum_{t=i+1}^{T}(x_t-\bar{x})\,e_t\,e_{t-l}(x_{t-l}-\bar{x})\}$$

To compute $\mathbf{S}_T$, we only add **12** autocovariances of w_t $(= x_t\, e_t)$ to the White estimator, $\mathbf{S}_0$.

Review – GRM: The Newey-West Estimator

• NW SEs are used almost universally in academia. However:
- NW SEs perform poorly in Monte Carlo simulations:
- NW SEs tend to be downward biased –i.e., too small.
- The finite-sample performance of tests using NW SE is not well approximated by the asymptotic theory.
- Tests have size distortions.

• Q: What happens if we know the specific form of (**A3'**)?
We can do much better than using OLS with NW SEs. In this case, we can do Generalized LS (GLS), a method that delivers the most efficient estimators.

Generalized Least Squares (GLS)

• GRM: Assumptions (**A1**), (**A2**), (**A3'**) & (**A4**) hold. That is,
(**A1**) DGP: $\boldsymbol{y} = \mathbf{X}\,\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ is correctly specified.
(**A2**) $\mathrm{E}[\boldsymbol{\varepsilon}|\mathbf{X}] = \mathbf{0}$
(**A3'**) $\mathrm{Var}[\boldsymbol{\varepsilon}|\mathbf{X}] = \boldsymbol{\Sigma} = \sigma^2\boldsymbol{\Omega}$ ($\boldsymbol{\Omega}$ is symmetric $\Rightarrow \mathbf{T}'\mathbf{T} = \boldsymbol{\Omega}$)
(**A4**) $\mathbf{X}$ has full column rank –i.e., rank($\mathbf{X}$) = k–, where T ≥ k.

• Suppose we know the form of (**A3'**)? We can use this information to gain efficiency.

• When we know (**A3'**), we transform $\boldsymbol{y}$ & $\mathbf{X}$, in such a way, that we can do again OLS with the transformed data.

To do this transformation, we exploit a property of symmetric matrices, like the variance-covariance matrix, $\boldsymbol{\Omega}$:

$$\boldsymbol{\Omega} \text{ is symmetric} \quad \Rightarrow \text{exists } \mathbf{T} \ni \quad \mathbf{T}'\mathbf{T} = \boldsymbol{\Omega} \quad \Rightarrow \mathbf{T}'^{-1}\,\boldsymbol{\Omega}\,\mathbf{T}^{-1} = \mathbf{I}$$

Generalized Least Squares (GLS)

Note: $\mathbf{\Omega}$ can be decomposed as

$$\mathbf{\Omega} = \mathbf{T}'\mathbf{T} \quad \text{(think of } \mathbf{T} \text{ as } \mathbf{\Omega}^{1/2}) \qquad \Rightarrow \mathbf{T}'^{-1}\,\mathbf{\Omega}\,\mathbf{T}^{-1} = \mathbf{I}$$

- We transform the linear model in (**A1**) using $\mathbf{P} = \mathbf{\Omega}^{-1/2}$ (= $\mathbf{T}^{-1}$).

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

$$\mathbf{P}\boldsymbol{y} = \mathbf{P}\mathbf{X}\boldsymbol{\beta} + \mathbf{P}\boldsymbol{\varepsilon} \quad \text{or}$$

$$\boldsymbol{y}^* = \boldsymbol{X}^*\boldsymbol{\beta} + \boldsymbol{\varepsilon}^*.$$

$$E[\boldsymbol{\varepsilon}^*\boldsymbol{\varepsilon}^{*\prime}|\boldsymbol{X}^*] = E[\mathbf{P}\boldsymbol{\varepsilon}\,\boldsymbol{\varepsilon}'\mathbf{P}'|\boldsymbol{X}^*] = \mathbf{P}\,E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|\mathbf{X}]\,\mathbf{P}' = \sigma^2\,\mathbf{P}\,\mathbf{\Omega}\,\mathbf{P}'$$

$$= \sigma^2\,\mathbf{\Omega}^{-1/2}\,\mathbf{\Omega}\,\mathbf{\Omega}^{-1/2} = \sigma^2\,\mathbf{I}_T \qquad \Rightarrow \text{back to (A3)}$$

- The transformed model is homoscedastic: We have the CLM framework back $\quad\Rightarrow$ we can use OLS!

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

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

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

Generalized Least Squares (GLS)

Remarks:

– The transformed model is homoscedastic:

$$\text{Var}[\boldsymbol{\varepsilon}^*|\boldsymbol{X}^*] = E[\boldsymbol{\varepsilon}^*\boldsymbol{\varepsilon}^{*\prime}|\boldsymbol{X}^*] = \mathbf{P}E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|\mathbf{X}]\mathbf{P}' = \sigma^2\,\mathbf{P}\mathbf{\Omega}\mathbf{P}' = \sigma^2\,\mathbf{I}_T$$

– We have the CLM framework back: We do OLS with the transformed model, we call this OLS estimator, the GLS estimator:

$$\mathbf{b}_{GLS} = \mathbf{b}^* = (\boldsymbol{X}^{*\prime}\boldsymbol{X}^*)^{-1}\boldsymbol{X}^{*\prime}\,\boldsymbol{y}^* = (\mathbf{X}'\mathbf{P}'\,\mathbf{P}\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{P}'\,\mathbf{P}\boldsymbol{y}$$

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

– Key assumption: $\mathbf{\Omega}$ is known, and, thus, $\mathbf{P}$ is also known; otherwise we cannot transformed the model.

- Big Question: Is $\mathbf{\Omega}$ known?

Alexander C. Aitken (1895 –1967, NZ)

Generalized Least Squares (GLS)

• The GLS estimator is:

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

Note I: $\mathbf{b}_{GLS} \neq \mathbf{b}$. $\mathbf{b}_{GLS}$ is BLUE by construction, $\mathbf{b}$ is not.

• Check unbiasedness:

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

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

$$E[\mathbf{b}_{GLS}\,|\mathbf{X}] = \boldsymbol{\beta}$$

• Efficient Variance

$\mathbf{b}_{GLS}$ is BLUE. The "best" variance can be derived from

$$\text{Var}[\mathbf{b}_{GLS}|\mathbf{X}] = \sigma^2\,(\mathbf{X}^{*\prime}\mathbf{X}^{*})^{-1} = \sigma^2\,(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}$$

Then, the usual OLS variance for **b** is biased and inefficient!

Generalized Least Squares (GLS) - Properties

Note II: Both unbiased and consistent. In practice, both estimators will be different, but not that different. If they are very different, something is not kosher.

• Steps for GLS:

Step 1. Find transformation matrix $\mathbf{P} = \boldsymbol{\Omega}^{-1/2}$ (in the case of heteroscedasticity, $\mathbf{P}$ is a diagonal matrix).

Step 2. Transform the model: $\boldsymbol{X}^{*} = \mathbf{P}\mathbf{X}$ & $\boldsymbol{y}^{*} = \mathbf{P}\boldsymbol{y}$.

Step 3. Do GLS; that is, OLS with the transformed variables.

• Key step to do GLS: **Step 1**, getting the transformation matrix:

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

GLS – Relaxing Assumptions (A2) & (A4)

Technical detail: If we relax the CLM assumptions (**A2**) and (**A4**), as we did in Lecture 7-a, we only have asymptotic properties for GLS:

– Consistency - "*well behaved data*."

– Asymptotic distribution under usual assumptions.
(easy for heteroscedasticity, complicated for autocorrelation.)

– Wald tests and *F*-tests with usual asymptotic χ^2 distributions.

(Weighted) GLS: Pure Heteroscedasticity

- Find the transformation matrix $\mathbf{P} = \mathbf{\Omega}^{-1/2}$.

$$(\mathbf{A3'})\quad Var[\varepsilon] = \mathbf{\Sigma} = \sigma^2\mathbf{\Omega} = \sigma^2\begin{bmatrix}\omega_1 & 0 & \dots & 0\\ 0 & \omega_2 & \dots & 0\\ 0 & 0 & & 0\\ 0 & 0 & \dots & \omega_T\end{bmatrix}$$

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

- Now, transform $\boldsymbol{y}$ & $\mathbf{X}$:

$$\boldsymbol{y}^* = \mathbf{P}\boldsymbol{y} = \begin{bmatrix}1/\sqrt{\omega_1} & 0 & \dots & 0\\ 0 & 1/\sqrt{\omega_2} & \dots & 0\\ 0 & 0 & \dots & 0\\ 0 & 0 & \dots & 1/\sqrt{\omega_T}\end{bmatrix} * \begin{bmatrix}y_1\\ y_2\\ \vdots\\ y_T\end{bmatrix} = \begin{bmatrix}y_1/\sqrt{\omega_1}\\ y_2/\sqrt{\omega_2}\\ \vdots\\ y_T/\sqrt{\omega_T}\end{bmatrix}$$

(Weighted) GLS: Pure Heteroscedasticity

- Each observation of y, y_i, is divided by $\sqrt{\omega_i}$. Similar transformation occurs with **X**:

$$\boldsymbol{X}^* = \mathbf{PX} = \begin{bmatrix} 1/\sqrt{\omega_1} & 0 & \dots & 0 \\ 0 & 1/\sqrt{\omega_2} & \dots & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 1/\sqrt{\omega_T} \end{bmatrix} * \begin{bmatrix} 1 & x_{21} & \cdots & x_{k1} \\ 1 & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & x_{2T} & \cdots & x_{kT} \end{bmatrix} =$$

$$= \begin{bmatrix} 1/\sqrt{\omega_1} & x_{21}/\sqrt{\omega_1} & \dots & x_{k1}/\sqrt{\omega_1} \\ 1/\sqrt{\omega_2} & x_{22}/\sqrt{\omega_2} & \dots & x_{k2}/\sqrt{\omega_2} \\ \vdots & \vdots & \dots & \vdots \\ 1/\sqrt{\omega_T} & x_{2T}/\sqrt{\omega_T} & \dots & x_{kT}/\sqrt{\omega_T} \end{bmatrix}$$

- Now, we can do OLS with the transformed variables:

$$\mathbf{b}_{\text{GLS}} = \mathbf{b}^* = (\boldsymbol{X}^{*\prime}\boldsymbol{X}^*)^{-1}\boldsymbol{X}^{*\prime}\, \boldsymbol{y}^* = (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\, \mathbf{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{y}$$

(Weighted) GLS: Pure Heteroscedasticity

- In the case of heteroscedasticity, GLS is also called *Weighted Least Squares* (WLS): Think of $1/\sqrt{\omega_i}$ as weights. The GLS estimator is:

$$\mathbf{b}_{\text{GLS}} = (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\, \mathbf{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{y} = \left(\sum_{i=1}^{T} \frac{1}{\omega_i} \boldsymbol{x}_i \boldsymbol{x}_i'\right)^{-1} \sum_{i=1}^{T} \frac{1}{\omega_i} \boldsymbol{x}_i y_i$$

Observations with lower (bigger) variances –i.e., lower (bigger) ω_i– are given higher (lower) weights in the sums:

More precise observations, more weight!

- The GLS variance is given by:

$$\hat{\sigma}^2_{GLS} = \frac{\sum_{i=1}^{T} \left(\frac{y_i - \boldsymbol{x}_i'\, \mathbf{b}_{\text{GLS}}}{\omega_i}\right)^2}{T-k}$$

(Weighted) GLS: Pure Heteroscedasticity

Example: Last Lecture, we found that squared market returns (Mkt_RF^2) influence the heteroscedasticity in DIS returns. Suppose we assume: (**A3'**) $\sigma_i^2 = (r_{m,t} - r_f)^2$.

Steps for GLS:

1. Find transformation matrix, **P**, with i^{th} diagonal element: $1/\sqrt{\sigma_i^2}$
2. Transform model: Each y_i and x_{i} is divided ("weighted") by $\sigma_i = \text{sqrt}[(r_{m,t} - r_f)^2]$.
3. Do GLS, that is, OLS with transformed variables.

```
T <- length(dis_x)
Mkt_RF2 <- Mkt_RF^2                                      # (A3')
y_w <- dis_x/sqrt(Mkt_RF2)                               # transformed y = y*
x0 <- matrix(1,T,1)
xx_w <- cbind(x0, Mkt_RF, SMB, HML)/sqrt(Mkt_RF2)        # transformed X = X*
fit_dis_wls <- lm(y_w ~ xx_w - 1)                        # GLS
```

(Weighted) GLS: Pure Heteroscedasticity

Example (continue):

```
> summary(fit_dis_wls)

Call:
lm(formula = y_w ~ xx_w)

Residuals:
    Min      1Q  Median      3Q     Max
-59.399  -0.891   0.316   1.503  77.434

Coefficients:
              Estimate  Std. Error t value Pr(>|t|)
xx_w          -0.006607  0.001586  -4.165 3.59e-05 ***
xx_wMkt_RF     1.588057  0.334771   4.744 2.66e-06 ***     ⇒ OLS b: 1.26056
xx_wSMB       -0.200423  0.067498  -2.969  0.00311 **      ⇒ OLS b: -0.028993
xx_wHML       -0.042032  0.072821  -0.577  0.56404         ⇒ OLS b:  0.174545
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.984 on 566 degrees of freedom
Multiple R-squared:  0.09078,   Adjusted R-squared:  0.08435
F-statistic: 14.13 on 4 and 566 DF,  p-value: 5.366e-11
```


GLS: First-order Autocorrelation Case

• We assume an AR(1) process for the ε_t:

$$\varepsilon_t = \rho\, \varepsilon_{t-1} + u_t, \qquad u_t\text{: } \textit{uncorrelated error} \text{ (WN)} \sim D(0, \sigma_u^2)$$

• We need to find the transformation matrix $\mathbf{P} = \mathbf{\Omega}^{-1/2}$ for:

$$(\mathbf{A3'})\;\; Var[\boldsymbol{\varepsilon}] = \boldsymbol{\Sigma} = \sigma^2\boldsymbol{\Omega} = \sigma^2 \begin{bmatrix} \omega & \omega_{12} & \cdots & \omega_{1T} \\ \omega_{21} & \omega & \cdots & \omega_{2T} \\ \vdots & \vdots & & \vdots \\ \omega_{T1} & \omega_{T2} & \cdots & \omega \end{bmatrix},$$

where $\sigma^2\omega_{ik} = \sigma_{ij} = E[\varepsilon_i\, \varepsilon_j] = \sigma^2\omega_{ij}$. $\quad (= \gamma_{l=i-j})$

$\sigma^2\, \omega = \sigma_{ii} = Var[\varepsilon_i]$ (homoscedasticity, constant for all *i*.)

<u>Notation</u>: We use γ_l to denote a (auto-) *covariance* between two observations separated by l periods. For example, when l =1:

$$\gamma_1 = \sigma_{21} = \sigma_{32} = \ldots = \sigma_{T(T-1)} = Cov[\varepsilon_t, \varepsilon_{t-1}] = E[\varepsilon_t\, \varepsilon_{t-1}]$$

GLS: First-order Autocorrelation Case

• Steps for GLS:

1. To find the transformation matrix **P**, we need to derive the implied (**A3'**) based on the AR(1) process for ε_t:

(1) Find diagonal elements of $\mathbf{\Omega}$: $Var[\varepsilon_t] = \sigma_{tt} = \sigma_\varepsilon^2$

$$\varepsilon_t = \rho\, \varepsilon_{t-1} + u_t \qquad \text{-autoregressive AR(1) form.}$$

We take variances on both sides –i.e., $E[\varepsilon_t^2]$:

$$Var[\varepsilon_t] = \rho^2\, Var[\varepsilon_{t-1}] + Var[u_t] \qquad (Var[\varepsilon_t] = Var[\varepsilon_{t-1}] = \sigma_\varepsilon^2)$$

$$\Rightarrow\;\; \sigma_\varepsilon^2 = \frac{\sigma_u^2}{(1-\rho^2)} \qquad \text{–we need to assume } |\rho| < 1.$$

<u>Notation</u>: Using the γ_l time series notation for (auto-) covariances:

$$\gamma_0 = \sigma_i^2 = E[\varepsilon_t\, \varepsilon_t] = Var[\varepsilon_t] = \sigma_u^2/(1-\rho^2).$$

$$\gamma_l = \sigma_{ij} = Cov[\varepsilon_i, \varepsilon_j] = E[\varepsilon_i\, \varepsilon_j] \qquad l = i - j$$

GLS: AR(1) Case – Autocovariances

(2) Find off-diagonal elements $\sigma_{ij} = \gamma_l$ (*autocovariance* γ_l at lag $l = j - i$). It measures how two errors separated in time by l *periods* covary:

$$\sigma_{ij} = \gamma_l = \text{Cov}[\varepsilon_i, \varepsilon_j] = \text{E}[\varepsilon_i\, \varepsilon_j] \qquad\qquad l = i - j$$

$$\gamma_1 = Cov[\varepsilon_t, \varepsilon_{t-1}] = E[(\rho\varepsilon_{t-1} + u_t)\, \varepsilon_{t-1}]$$
$$= \rho\, E[\varepsilon_{t-1}\, \varepsilon_{t-1}] + E[u_t\, \varepsilon_{t-1}]$$
$$= \rho\, \text{Var}[\varepsilon_{t-1}] = \rho\, \sigma_\varepsilon^2$$
$$= \rho\, \gamma_0$$

$$\gamma_2 = Cov[\varepsilon_t, \varepsilon_{t-2}] = E[(\rho\varepsilon_{t-1} + u_t)\, \varepsilon_{t-2}]$$
$$= \rho\, E[\varepsilon_{t-1}\, \varepsilon_{t-2}] + E[u_t\, \varepsilon_{t-2}]$$
$$= \rho\, Cov[\varepsilon_t, \varepsilon_{t-1}]$$
$$= \rho\, \gamma_1$$
$$= \rho^2 \gamma_0$$

GLS: AR(1) Case – Autocovariances

$$\gamma_2 = \rho\, \gamma_1 = \rho^2 \gamma_0$$

$$\gamma_3 = Cov[\varepsilon_t, \varepsilon_{t-3}] = E[(\rho\varepsilon_{t-1} + u_t)\, \varepsilon_{t-3}]$$
$$= \rho\, E[\varepsilon_{t-1}\, \varepsilon_{t-3}] + E[u_t\, \varepsilon_{t-3}]$$
$$= \rho\, Cov[\varepsilon_t, \varepsilon_{t-2}] = \rho\, \gamma_2$$
$$= \rho^2 \gamma_1$$
$$= \rho^3\, \gamma_0$$

⋮

$$\gamma_l = Cov[\varepsilon_t, \varepsilon_{t-l}] = \rho^{l-1}\, \gamma_1 \qquad\qquad (\gamma_1 = \rho\, \gamma_0)$$
$$= \rho^l\, \gamma_0$$

- We defined: $\gamma_0 = \sigma_\varepsilon^2 = \frac{\sigma_u^2}{(1-\rho^2)}$,

$\Rightarrow$ $\gamma_l = \rho^l \frac{\sigma_u^2}{(1-\rho^2)}$

GLS: AR(1) Case – Autocorrelation Matrix Σ

- Now, we get (**A3'**) $\boldsymbol{\Sigma} = \sigma^2\,\boldsymbol{\Omega}$.

$$(\textbf{A3'})\quad \sigma^2\Omega = \left(\frac{\sigma_u^2}{1-\rho^2}\right)\begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{T-1} \\ \rho & 1 & \rho & \cdots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \cdots & 1 \end{bmatrix}$$

- Then, we can get the transformation matrix $\mathbf{P} = \boldsymbol{\Omega}^{-1/2}$:

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

GLS: AR(1) Case – Transformed $\boldsymbol{y}$ & X: $\boldsymbol{y}^*$ & X*

2. With $\mathbf{P} = \boldsymbol{\Omega}^{-1/2}$, we transform the data to do GLS.

$$\mathbf{P} = \boldsymbol{\Omega}^{-1/2} = \begin{bmatrix} \sqrt{1-\rho^2} & 0 & 0 & \dots & 0 \\ -\rho & 1 & 0 & \dots & 0 \\ 0 & -\rho & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & -\rho & 0 \end{bmatrix} * \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_T \end{bmatrix}$$

:

$$\boldsymbol{y}^* = \mathbf{P}\,\boldsymbol{y} = \begin{pmatrix} \left(\sqrt{1-\rho^2}\right) y_1 \\ y_2 - \rho\, y_1 \\ y_3 - \rho\, y_2 \\ \dots \\ y_T - \rho\, y_{T-1} \end{pmatrix}$$

⇒ GLS: Transformed $\boldsymbol{y}^*$.

GLS: AR(1) Case – Transformed $\boldsymbol{y}$ & X: $\boldsymbol{y}^*$ & X*

2. Transformed $\boldsymbol{x}_k$ column (independent variable k) of matrix **X** is:

$$\mathbf{P} = \boldsymbol{\Omega}^{-1/2} = \begin{bmatrix} \sqrt{1-\rho^2} & 0 & 0 & \dots & 0 \\ -\rho & 1 & 0 & \dots & 0 \\ 0 & -\rho & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & -\rho & 0 \end{bmatrix} * \begin{bmatrix} x_{k1} \\ x_{k2} \\ x_{k3} \\ \vdots \\ x_{kT} \end{bmatrix}$$

$$\boldsymbol{x}_k^* = \mathbf{P}\,\boldsymbol{x}_k = \begin{pmatrix} \left(\sqrt{1-\rho^2}\right) x_{k1} \\ x_{k2} - \rho\, x_{k1} \\ x_{k3} - \rho\, x_{k2} \\ \dots \\ x_T - \rho\, x_{T-1} \end{pmatrix}$$

$\Rightarrow$ GLS: Transformed $\boldsymbol{X}^*$.

3. GLS is done with transformed data. In (**A3'**) we assume ρ known.

GLS: The Autoregressive Transformation

- With AR models, sometimes it is easier to transform the data by taking *pseudo differences.*

- For the AR(1) model, we multiply the DGP by ρ and subtract it from it. That is,

$$y_t = \boldsymbol{x}_t{}'\boldsymbol{\beta} + \varepsilon_t, \qquad \varepsilon_t = \rho\varepsilon_{t-1} + u_t$$
$$\rho y_{t-1} = \rho \boldsymbol{x}_{t-1}{}'\boldsymbol{\beta} + \rho\varepsilon_{t-1}$$
$$\text{————————————————————— –}$$
$$y_t - \rho y_{t-1} = (\boldsymbol{x}_t - \rho\boldsymbol{x}_{t-1})'\boldsymbol{\beta} + (\varepsilon_t - \rho\varepsilon_{t-1})$$
$$y_t^* = \boldsymbol{x}_t^*{}'\boldsymbol{\beta} + u_t$$

Now, we have the errors, u_t, which are uncorrelated. We can do OLS with the pseudo differences.

Note: $y_t^* = y_t - \rho y_{t-1}$ & $\boldsymbol{x}_t^* = \boldsymbol{x}_t - \rho\boldsymbol{x}_{t-1}$ are *pseudo differences.*

FGLS: Unknown Ω

• The problem with GLS is that $\boldsymbol{\Omega}$ is unknown. For example, in the AR(1) case, ρ is unknown.

• Solution: Estimate $\boldsymbol{\Omega}$. $\Rightarrow$ *Feasible GLS* (FGLS).

• In general, there are two approaches for GLS:

(1) Two-step, or *Feasible estimation*:
- First, estimate $\boldsymbol{\Omega}$ first.
- Second, do GLS.

Technical note: Nice asymptotic properties for FGLS estimator. Not longer BLUE.

(2) ML estimation of $\boldsymbol{\beta}$, σ^2, and $\boldsymbol{\Omega}$ at the same time (joint estimation of all parameters). With some exceptions, rare in practice.

FGLS: Specification of Ω

• $\boldsymbol{\Omega}$ must be specified first.

• In general, $\boldsymbol{\Omega}$ is specified in terms of a few parameters. Thus, $\boldsymbol{\Omega} = \boldsymbol{\Omega}(\boldsymbol{\theta})$ for some small parameter vector $\boldsymbol{\theta}$. Then, we need to estimate $\boldsymbol{\theta}$.

Examples:

(1) $\text{Var}[\varepsilon_i|\mathbf{X}] = \sigma^2 f(\boldsymbol{\gamma}'\mathbf{z}_i)$. Variance a function of $\boldsymbol{\gamma}$ and some variable $\mathbf{z}_i$ (say, market volatility, firm size, industry dummy, etc). In general, f(.) is an exponential to make sure the variance is positive.

(2) ε_i with AR(1) process. We have already derived $\sigma^2\,\boldsymbol{\Omega}$ as a function of ρ.

Technical note: To achieve full efficiency, we do not need an *efficient* estimate of the parameters in $\boldsymbol{\Omega}$, only a *consistent* one.

FGLS: Estimation – Steps

- Steps for FGLS:

1. Estimate the model proposed in (**A3'**). Get $\hat{\sigma}_i^2$ & $\hat{\sigma}_{ij}$.
2. Find transformation matrix, **P**, using the estimated $\hat{\sigma}_i^2$ & $\hat{\sigma}_{ij}$.
3. Using **P** from Step 2, transform model: $\boldsymbol{X}^* = \mathbf{P}\mathbf{X}$
   $\boldsymbol{y}^* = \mathbf{P}\boldsymbol{y}$.
4. Do FGLS, that is, OLS with $\boldsymbol{X}^*$ & $\boldsymbol{y}^*$.

Example: In the pure heteroscedasticity case (**P** is diagonal):
1. Estimate the model proposed in (**A3'**). Get $\hat{\sigma}_i^2$.
2. Find transformation matrix, **P**, with i^{th} diagonal element: $1/\hat{\sigma}_i$
3. Transform model (each y_i and x_i is divided ("weighted") by $\hat{\sigma}_i$):

$$y_i^* = y_i/\hat{\sigma}_i$$
$$x_{k,i}^* = x_{k,i}/\hat{\sigma}_i$$

4. Do FGLS, that is, OLS with transformed variables.

FGLS: Estimation – Heteroscedasticity

Example: Last lecture, we found that $(r_{m,t} - r_f)^2$ & $(SMB_t)^2$ are drivers of the heteroscedasticity in DIS returns: Suppose we assume:
(**A3'**) $\sigma_t^2 = \gamma_0 + \gamma_1\,(r_{m,t} - r_f)^2 + \gamma_3\,(SMB_t)^2$

- Steps for FGLS:

1. Use OLS squared residuals to estimate (**A3'**):

```
fit_dis_ff3 <- lm(dis_x ~ Mkt_RF + SMB + HML)
e_dis <- fit_dis_ff3$residuals
e_dis2 <- e_dis^2
fit_dis2 <- lm(e_dis2 ~ Mkt_RF2 + SMB2)
summary(fit_dis2)
var_dis2 <- fit_dis2$fitted          # Estimated variance vector, with elements σ̂_i^2.
```

2. Find transformation matrix, **P**, with i^{th} diagonal element: $1/\hat{\sigma}_i$

```
w_fgls <- sqrt(var_dis2)             # 1/σ̂_i
```

3. Transform model: Each y_i and x_i is "weighted" by $1/\hat{\sigma}_i$.

```
y_fw <- dis_x/w_fgls                             # transformed y
xx_fw <- cbind(x0, Mkt_RF, SMB, HML)/w_fgls      # transformed X
```


FGLS: Estimation – Heteroscedasticity

Example (continuation):
4. Do GLS, that is, OLS with transformed variables.

```
fit_dis_fgls <- lm(y_fw ~ xx_fw - 1)
> summary(fit_dis_fgls)

Coefficients:
                Estimate   Std. Error t value Pr(>|t|)
xx_fw           -0.003097  0.002696  -1.149   0.251
xx_fwMkt_RF     1.208067   0.073344  16.471   <2e-16 ***
xx_fwSMB        -0.043761  0.105280  -0.416   0.678
xx_fwHML        0.125125   0.100853   1.241   0.215      ⇒ not longer significant at 10%.
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9998 on 566 degrees of freedom
Multiple R-squared:  0.3413,    Adjusted R-squared:  0.3366
F-statistic: 73.31 on 4 and 566 DF,  p-value: < 2.2e-16
```

FGLS: Estimation – Heteroscedasticity

Example (continuation): Comparing OLS, GLS & FGLS results:

	b_{OLS}	SE	b_{GLS}	SE	b_{FGLS}	SE
Intercept	0.00417	0.00279	-0.00661	0.00159	-0.00310	0.00270
Mkt_RF	**1.26056**	0.06380	**1.58806**	0.33477	**1.20807**	0.07334
SMB	-0.02899	0.09461	-0.20042	0.06750	-0.04376	0.10528
HML	0.17455	0.09444	-0.04203	0.07282	0.12513	0.10085

• Comments:
- The GLS estimates are quite different than OLS estimates (remember OLS is unbiased and consistent). Very likely the assumed functional form in (**A3'**) was not a good one.
- The FGLS results are similar to the OLS, as expected, if model is OK. FGLS is likely a more precise estimator (HML is not longer significant at 10%.

FGLS Estimation: AR(1) Case – Cochrane-Orcutt

- In the AR(1) case, it is easier to estimate the model in *pseudo differences*:

$$y_t^* = \mathbf{X}_t^* \beta + u_t$$

$$y_t - \rho y_{t-1} = (\mathbf{X}_t - \rho \mathbf{X}_{t-1})' \beta + \varepsilon_t - \rho\, \varepsilon_{t-1}$$

$$\Rightarrow y_t = \rho y_{t-1} + \mathbf{X}_t' \beta - \mathbf{X}_{t-1}' \rho\beta + u_t$$

- We have a linear model, but it is nonlinear in parameters. OLS is not possible, but non-linear estimation is possible.

- Before today's computer power, Cochrane–Orcutt's (1949) iterative procedure was an ingenious way to do this estimation.

FGLS Estimation: AR(1) Case – Cochrane-Orcutt

- Steps for Cochrane-Orcutt:

(0) Do OLS in (**A1**) model: $\boldsymbol{y} = \mathbf{X}\, \boldsymbol{\beta} + \boldsymbol{\varepsilon}$. Get residuals, $\boldsymbol{e}$, & RSS_0.

(1) Estimate ρ with a regression of $\boldsymbol{e}_t$ against $\boldsymbol{e}_{t-1}$ $\Rightarrow$ get $\hat{\rho}$ (the estimator of ρ).

(2) FGLS Step. Use $\hat{\rho}$ transform the model to get $\boldsymbol{y}^*$ and $\mathbf{X}^*$.

Do OLS with $\boldsymbol{y}^*$ and $\boldsymbol{X}^*$ $\Rightarrow$ get **b** to estimate β.

Get residuals, $\boldsymbol{e}^* = \boldsymbol{y} - \mathbf{X}\, \mathbf{b}$, and new RSS_1. Go back to (1).

(3) Iterate until convergence, usually achieved when the difference in RSS of two consecutive iterations is lower than some tolerance level, say .0001. Then, stop when $RSS_i - RSS_{i-1} < .0001$.

FGLS Estimation: Cochrane-Orcutt in R

Example: Cochrane-Orcutt in R

```
# C.O. function requires Y, X (with constant), OLS b.
c.o.proc <- function(Y,X,b_0,tol){
 T <- length(Y)
 e <- Y - X%*%b_0                      # OLS residuals
rss <- sum(e^2)                        # Initial RSS of model, RSS9
rss_1 <- rss                           # RSS_1 will be used to reset RSS after each iteration
d_rss = rss                            # initialize d_rss: difference between RSSi & RSSi-1
 e2 <- e[-1]                           # adjust sample size for et
 e3 <- e[-T]                           # adjust sample size for et-1
 ols_e0 <- lm(e2 ~ e3 - 1)             # OLS to estimate rho
 rho <- ols_e0$coeff[1]                # initial value for rho, ρ0
 i<-1
while (d_rss > tol) {                  # tolerance of do loop. Stop when diff in RSS < tol
rss <- rss_1                           # RSS at iter (i-1)
  YY <- Y[2:T] - rho * Y[1:(T-1)]      # pseudo-diff Y
  XX <- X[2:T, ] - rho * X[1:(T-1), ]  # pseudo-diff X
ols_yx <- lm(YY ~ XX - 1)              # adjust if constant included in X
```

FGLS Estimation: Cochrane-Orcutt in R

Example (continuation):

```
b <- ols_yx$coef                       # updated OLS b at iteration i
#  b[1] <- b[1]/(1-rho)                # If constant not pseudo-differenced remove tag #
e1 <- Y - X%*%b                        # updated residuals at iteration i
   e2 <- e1[-1]                        # adjust sample size for updated et
   e3 <- e1[-T]                        # adjust sample size for updated e_t-1 (lagged et)
   ols_e1 <- lm(e2~e3-1)               # updated regression to value for rho at iteration i
   rho <- ols_e1$coeff[1]              # updated value of rho at iteration i, ρi
 rss_1 <- sum(e1^2)                    # updated value of RSS at iteration i, RSSi
 d_rss <- abs(rss_1 - rss)             # diff in RSS (RSSi - RSSi-1)
i <- i+1
   }

 result <-list()
 result$Cochrane_Orc.Proc <- summary(ols_yx)
 result$rho.regression <- summary(ols_e1)
 #  result$Corrected.b_1 <- b[1]
 result$Iterations < -i-1
 return(result)
}
```

FGLS Estimation: Cochrane-Orcutt – i_{MX}

Example: In the model for Mexican interest rates (i_{MX}), we suspect an AR(1) in the residuals:

$$i_{MX,t} = \beta_0 + \beta_1\, i_{US,t} + \beta_2\, e_t + \beta_3\, mx\_I_t + \beta_4\, mx\_y_t + \varepsilon_t$$
$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t$$

• **OLS estimation**.

```
y <- mx_i_1
T_mx <- length(mx_i_1)
x0 <- matrix(1,T_mx,1)
X <- cbind(x0, us_i_1, e_mx, mx_I, mx_y)                    # X matrix
fit_i <- lm(mx_i_1 ~ us_i_1 + e_mx + mx_I + mx_y)
b_i <- fit_i$coefficients                                    # extract coefficients from lm
> summary(fit_i)
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.04022    0.01506   2.671  0.00834 **
us_i_1       0.85886    0.31211   2.752  0.00661 **
e_mx        -0.01064    0.02130  -0.499  0.61812
mx_I         3.34581    0.19439  17.212  < 2e-16 ***
mx_y        -0.49851    0.73717  -0.676  0.49985
```

FGLS Estimation: Cochrane-Orcutt – i_{MX}

Example (continuation): Now, we use **Cochrane-Orcutt**:

```
> c.o.proc(y, X, b_i, .0001)
$Cochrane.Orcutt.Proc

Call:
lm(formula = YY ~ XX - 1)

Residuals:
     Min       1Q   Median       3Q      Max
-0.69251 -0.02118 -0.01099  0.00538  0.49403

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
XX          0.16639    0.07289   2.283   0.0238 *
XXus_i_1    1.23038    0.76520   1.608   0.1098        ⇒ not longer significant at 5% level.
XXe_mx     -0.00535    0.01073  -0.499   0.6187
XXmx_I      0.41608    0.27260   1.526   0.1289        ⇒ not longer significant at 5% level.
XXmx_y     -0.44990    0.53096  -0.847   0.3981
---
```

FGLS Estimation: Cochrane-Orcutt – i_{MX}

Example (continuation):

```
Residual standard error: 0.09678 on 160 degrees of freedom
Multiple R-squared:  0.1082,    Adjusted R-squared:  0.08038
F-statistic: 3.884 on 5 and 160 DF,  p-value: 0.002381

$rho
      e3
0.8830857                                ⇒ very high autocorrelation.

$Corrected.b_1
      XX
0.1663884                        ⇒ Constant corrected if X does not include a constant

$Number.Iteractions
[1] 10                           ⇒ algorithm converged in 10 iterations.
```

Note: The R package "orcutt" computes the Cochrane-Orcutt algorithm:
library(orcutt)
cochrane.orcutt(**fit_i**, convergence = 8, max.iter=100)

GLS: General Remarks

- GLS is great (BLUE) if we know **Ω**. Very rare situation.
- It needs the specification of **Ω** –i.e., the functional form of autocorrelation and heteroscedasticity.
- If the specification is bad ⇒ estimates are biased.
- Feasible GLS is not BLUE (unlike GLS); but, it is consistent and asymptotically more efficient than OLS.
- We use GLS for inference and/or efficiency. OLS is still unbiased and consistent.
- OLS and GLS estimates will be different due to sampling error. But, if they are very different, then it is likely that some other CLM assumption is violated.