Lecture 7-d
GLS & FGLS

Brooks (4th edition): Chapter 5

 1

Review - Generalized Regression Model (GRM)

• Assumptions behind the generalized regression model (GRM): (**A1**) DGP: $\boldsymbol{y} = \mathbf{X}\,\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ is correctly specified.

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

(**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**) $\mathbf{X}$ has full column rank – rank($\mathbf{X}$) = k –, where $T \geq k$.

• OLS is still unbiased (& consistent). Can we still use OLS?
Yes! But, we need to make inferences based on White or NW SE.

Review – Generalized Least Squares (GLS)

• If we know the specific form of (**A3'**), we can do better that OLS with White/NE SE. We can gain efficiency using **GLS**.

• We transform (**A1**) using $\mathbf{P} = \mathbf{\Omega}^{-1/2}$ $\qquad (\Rightarrow \mathbf{P}'\mathbf{P} = \mathbf{\Omega}^{-1})$

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

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

Then,

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

Back to CLM with modified model:

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

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

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

• Key assumption: $\mathbf{\Omega}$ is known, and, thus, $\mathbf{P}$ is also known.

Review – Generalized Least Squares (GLS)

• The GLS estimator is:

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

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

Note II: Both unbiased & consistent. In practice, both estimators will be different, but not that different. If they are very different, worry.

• Steps for GLS:

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

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

Review – GLS: Pure Heteroscedasticity

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

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

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

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

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

Review – GLS: Pure Heteroscedasticity

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

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

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

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

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

Review – GLS: Pure Heteroscedasticity

• Note: In the heteroscedasticity case, GLS is also called **Weighted Least Squares** (**WLS**). Each observation receives a $1/\sqrt{\omega_i}$ weigth, inverse to the SD of the error ε_i.

More precise observations (lower $\sqrt{\omega_i}$), more weight!

Example: Last Lecture, we assumed: (**A3'**) $\sigma_i^2 = (r_{m,t} - r_f)^2$.

Then, $y_t^* = (r_{i,t} - r_f)/(r_{m,t} - r_f)^2$.

$\boldsymbol{X}_t^* = [1/(r_{m,t} - r_f)^2, (r_{m,t} - r_f)/(r_{m,t} - r_f)^2,$
$SMB_t\ /(r_{m,t} - r_f)^2, HML_t/(r_{m,t} - r_f)^2]$

And we did OLS with transformed data: y_t^* & $\boldsymbol{X}_t^*$.

Problem: Who knows (**A3'**) ? In general, we have a model for $\boldsymbol{\Sigma}$, which we estimate $\boldsymbol{\Sigma}$. Then, we do **FGLS**.

Review – GLS: AR(1) Case – Autocovariances

• 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 $\mathbf{P} = \boldsymbol{\Omega}^{-1/2}$ for:

$$\textbf{(A3')}\quad Var[\boldsymbol{\varepsilon}] = \boldsymbol{\Sigma} = \begin{bmatrix} \sigma^2 & \sigma_{12} & \cdots & \sigma_{1T} \\ \sigma_{21} & \sigma^2 & \cdots & \sigma_{2T} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{T1} & \sigma_{T2} & \cdots & \sigma^2 \end{bmatrix},$$

which we will decompose into $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{\Omega}$ (our goal: get $\mathbf{P} = \boldsymbol{\Omega}^{-1/2}$)

Review – GLS: First-order Autocorrelation Case

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

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

$l = 2$: $\gamma_2 = \sigma_{31} = \sigma_{42} = \ldots = \sigma_{T(T-2)} = \text{Cov}[\varepsilon_t, \varepsilon_{t-2}] = \text{E}[\varepsilon_t\, \varepsilon_{t-2}]$

γ_l measures how two errors separated in time by l *periods* covary

- Let $\gamma_0 = \sigma_\varepsilon^2 = \text{E}[\varepsilon_t\, \varepsilon_t]$. Then, we can write (**A3'**) as:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma^2 & \sigma_{12} & \cdots & \sigma_{1T} \\ \sigma_{21} & \sigma^2 & \cdots & \sigma_{2T} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{T1} & \sigma_{T2} & \cdots & \sigma^2 \end{bmatrix} = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{T-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{T-2} \\ \vdots & \vdots & \vdots & \vdots \\ \gamma_{T-1} & \gamma_{T-2} & \cdots & \gamma_0 \end{bmatrix}.$$

Remark: Eventually, decompose $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{\Omega}$, since we need $\mathbf{P} = \boldsymbol{\Omega}^{-1/2}$.

Review – GLS: AR(1) Case

- Steps for GLS:

Step 1. Find the transformation matrix **P**.

We need to derive $\boldsymbol{\Sigma}$, based on the AR(1) process for ε_t:

(1) Find diagonal elements of Ω:

$$\gamma_0 = \text{Var}[\varepsilon_t] = \sigma_\varepsilon^2$$

Using the AR(1) model: $\varepsilon_t = \rho\, \varepsilon_{t-1} + u_t$.

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

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

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

Now, we have all the diagonal elements of $\boldsymbol{\Sigma}$.

Review – GLS: AR(1) Case

(2) Find off-diagonal elements of Ω:

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

$\gamma_1 = Cov[\varepsilon_t, \varepsilon_{t-1}] = \rho\, \gamma_0$

$\gamma_2 = Cov[\varepsilon_t, \varepsilon_{t-2}] = \rho^2 \gamma_0$

$\vdots$

$\gamma_l = Cov[\varepsilon_t, \varepsilon_{t-l}] = \rho^l\, \gamma_0$

Then,

$$\boldsymbol{\Sigma} = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{T-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{T-2} \\ \vdots & \vdots & \vdots & \vdots \\ \gamma_{T-1} & \gamma_{T-2} & \cdots & \gamma_0 \end{bmatrix} = \begin{bmatrix} \gamma_0 & \rho\gamma_0 & \cdots & \rho^{T-1}\gamma_0 \\ \rho\gamma_0 & \gamma_0 & \cdots & \rho^{T-2}\gamma_0 \\ \vdots & \vdots & \vdots & \vdots \\ \rho^{T-1}\gamma_0 & \rho^{T-2}\gamma_0 & \cdots & \gamma_0 \end{bmatrix}$$

Note: We take γ_0 out of the matrix. It becomes σ^2 in the $\boldsymbol{\Sigma}$ into $\sigma^2\boldsymbol{\Omega}$

Review – GLS: AR(1) Case – Matrix Σ

- We defined $\gamma_0 = \sigma_\varepsilon^2 = \frac{\sigma_u^2}{(1-\rho^2)}$. Then, decompose $\boldsymbol{\Sigma}$ into $\sigma^2\boldsymbol{\Omega}$.

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

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

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

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

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

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

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

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

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

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

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

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

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

Step 3. Do GLS: OLS with transformed data. (Q: Is ρ known?).

Review – GLS: The AR Transformation

• Easier derivation for AR models. For example, for the AR(1) model, we multiply the DGP by ρ and subtract it from it:

$$y_t = \boldsymbol{x}_t'\boldsymbol{\beta} + \varepsilon_t, \qquad \varepsilon_t = \rho\varepsilon_{t-1} + u_t$$

$$\rho y_{t-1} = \rho\boldsymbol{x}_{t-1}'\boldsymbol{\beta} + \rho\varepsilon_{t-1}$$

$$\text{———————————————————— –}$$

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

$$y_t^* = \boldsymbol{x}_t^{*\prime}\boldsymbol{\beta} + u_t$$

Now, the errors, u_t, 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*.

Review – FGLS: Unknown $\boldsymbol{\Omega}$

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

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

• To do the estimation, $\boldsymbol{\Omega}$ must be specified first. Usually, as a function $\boldsymbol{\Omega} = \boldsymbol{\Omega}(\boldsymbol{\theta})$, for some small parameter vector $\boldsymbol{\theta}$.

• In general, two approaches for GLS estimation:

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

(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

Examples:

(1) Var[ε_i|**X**] = $\sigma^2 f(\gamma' z_i)$. Variance a function of **γ** and some variable $\boldsymbol{z}_i$ (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 σ^2 **Ω** as a function of ρ.

Technical note: Two-step estimation has nice asymptotic properties for FGLS estimator. But, FGLS is no longer BLUE.

Review – 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'**) $\sigma_t^2 = \gamma_0 + \gamma_1 (r_{m,t} - r_f)^2 + \gamma_2 (SMB_t)^2$

• 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 σ̂i².
```

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

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

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

• AR(1) case: It is easier to estimate the model in *pseudo differences*:

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

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

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

• OLS cannot estimate ρ & **β**. We need a non-linear estimation, like Cochrane–Orcutt's (1949) iterative procedure.

<u>Note</u>: We can do a regression:

$$y_t = \delta_1\, y_{t-1} + \mathbf{X}_t' \boldsymbol{\delta_2} - \mathbf{X}_{t-1}' \boldsymbol{\delta_3} + u_t$$

OLS will estimate δ_1, δ_2, & δ_3. To get ρ & **β**, we need a restriction:

$$\delta_1 * \boldsymbol{\delta_2} = -\, \boldsymbol{\delta_3}$$

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

- Steps for Cochrane-Orcutt:

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

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

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

Do OLS with $\boldsymbol{y}^*$ and $\boldsymbol{X}^*$ ⇒ 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$coef                      # updated OLS b at iteration i
# b[1] <- b[1]/(1-rho)                # If constant not pseudo-differenced remove tag
#
e1 <- Y - X%*%b                       # updated residuals at iteration i
  e2 <- e1[-1]                        # adjust sample size for updated e_t
  e3 <- e1[-T]                        # adjust sample size for updated e_t-1 (lagged e_t)
  ols_e1 <- lm(e2~e3-1)               # updated regression to value for rho at iteration i
  rho <- ols_e1$coeff[1]              # updated value of rho at iteration i, ρ̂_i
 rss_1 <- sum(e1^2)                   # updated value of RSS at iteration i, RSS_i
 d_rss <- abs(rss_1 - rss)            # diff in RSS (RSS_i - RSS_i-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)
xx_i <- cbind(us_i_1, e_mx, mx_I, mx_y)
x0 <- matrix(1,T_mx,1)
X <- cbind(x0,xx_i)                                    # 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)  0.02712   0.01265   2.144  0.03337 *
us_i_1       0.68262   0.24955   2.735  0.00687 **
e_mx        -0.01399   0.01851  -0.756  0.45078
mx_I         3.66118   0.15950  22.955  < 2e-16 ***
mx_y         0.04659   0.22505   0.207  0.83623
```

FGLS Estimation: Cochrane-Orcutt – i_{MX}

Example (continuation): Cochrane-Orcutt estimation, with orcutt package (should give same results as c.o.proc(y, X, b_i, .0001).

```
library(orcutt)
coch_i <- cochrane.orcutt(fit_i, convergence = 8, max.iter=100)
t_coch_i <- coch_i$t.value      # Extract t-values from coch_i

>coch_i
 number of interaction: 18
 rho 0.88476

 coefficients:
(Intercept)    us_i_1      e_mx       mx_I      mx_y
   0.133528  0.819452 -0.008638  1.261720  0.026916
> t_coch_i
(Intercept)    us_i_1      e_mx       mx_I      mx_y
  2.3040067  1.3425452 -0.9794369  5.7474954  0.2145424
```

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.
```

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 bad (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.