EC402 (2023/24)

RAGVIR'S SPEAKING NOTES FOR CLASS

Vassilis PS 5 (WK 6, AT)

QUESTION 1

Q1. (a) –

(b) under A4GM: $E[\varepsilon\varepsilon'] = \sigma^2 I_T$

So, $Var(\hat\beta_{LSE}) = E\left[\left(\hat\beta_{LSE} - E(\hat\beta_{LSE})\right)\left(\hat\beta_{LSE} - E(\hat\beta_{LSE})\right)'\right]$ due to ... ,

where $\hat\beta_{LSE} := (X'X)^{-1}X'y$;

$= E\left[(X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1}\right]$ due to ... ;

$= (X'X)^{-1}X' E[\varepsilon\varepsilon'] X(X'X)^{-1}$ due to ... ;

$= (X'X)^{-1}X'\sigma^2 I_T X(X'X)^{-1}$ due to ... ;

$= \sigma^2(X'X)^{-1}$.

But here, we do not have A4GM!

Under A4Ω: $E[\varepsilon\varepsilon'] = \Omega$, a p.d. symmetric matrix,

$Var(\hat\beta_{LSE}) = \ldots = (X'X)^{-1}X'\Omega X(X'X)^{-1}$

(c) · Of course not! Under A4 Ω, if $\Omega \neq \sigma^2 I_T$, then $\sigma^2(X'X)^{-1}$ is just some arbitrary matrix we came up with. It is utterly meaningless.

· Under the G/M theorem, $\hat{\beta}_{GLS} := (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$, is the BLUE.

That is, $\lambda_T\{ Var(\hat{\beta}_{LSE}) - Var(\hat{\beta}_{GLS}) \} > 0$, where $\lambda_T\{.\}$ refers to the smallest eigenvalue;

$$\text{or} \quad \lambda_T \{ (X'X)^{-1}X'\Omega^{-1}X(X'X)^{-1} - (X'\underline{\Omega^{-1}}X)^{-1} \} \geq 0 .$$

(i.e. so $\sigma^2(X'X)^{-1}$ really doesn't appear anywhere.)

Vassilis verifies this in his solutions.

make sure you can derive this.

QUESTION 2

Q2 (a) Say $y_t = \beta_0 + \sum_{h=1}^{K} \beta_h x_{ht} + \varepsilon_t$, for $t = 1, \ldots, T$.

UNDERGRAD STYLE

Then $y_{t-1} = \beta_0 + \sum_{h=1}^{K} \beta_h x_{ht-1} + \varepsilon_{t-1}$, for $t = 2, \ldots, T$.

So the "FD model" is defined as

$$\Delta y_t = \sum_{h=1}^{K} \beta_h \Delta x_{ht} + \Delta \varepsilon_t, \text{ for } t = 2, \ldots, T,$$

where $\Delta y_t := y_t - y_{t-1}$ and Δx_{ht} and $\Delta \varepsilon_t$ are defined analogously.

Say $y = i\beta_0 + X\beta + \varepsilon$, where

$$\begin{cases} X \text{ is a } T \times K \text{ matrix}; \\ i \text{ is a } T \text{ dimensional vector of ones}; \\ \varepsilon_t \overset{\text{IID}}{\sim} (0, \sigma_\varepsilon^2) \text{ for } t = 1, \ldots, T. \end{cases}$$

POSTGRAD STYLE

(styles are equivalent)

Then the "FD Model" is defined as Ay where A is ...

let's visualise together...

IGNORE

$$\begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \end{bmatrix}$$
$$\begin{bmatrix} 0 & -1 & 1 & \cdots & 0 \end{bmatrix}$$

DATA

y	i	x_1	$\cdots$
y_1	1	x_{11}	
y_2	1	x_{12}	
y_3	1	x_{13}	
$\vdots$	$\vdots$	$\vdots$	$\cdots$
y_{T-1}	1	x_{1T-1}	
y_T	1	x_{1T}	

NEW DATA

y	i	x_1
.	.	.
$y_2 - y_1$	0	$x_{12} - x_{11}$
$y_3 - y_2$	0	$x_{13} - x_{12}$

So if we define a $(T-1) \times T$ matrix

$$A := \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{pmatrix}$$

then the transformation $Ay = Ax\beta + A\varepsilon$, yields a "FD model" as required.

Note: $Ai = 0$. In fact, $A(ci) = 0$ for any constant $c \in \mathbb{R}$.

(b) Say the G/M assumptions held in the "levels model".

A1: $[i\ X]$ has rank $K+1$

A2: $y = [i\ x]\begin{bmatrix}\beta_0 \\ \beta\end{bmatrix} + \varepsilon$, $E[\varepsilon] = 0$

A3F: $[i\ x]$ is fixed in repeated samples

A4GM: $E[\varepsilon\varepsilon'] = \sigma^2 I_T$

- Then the FWL theorem and the G/M theorem tell us

$$\hat{\beta}_{OLS} = (X'M_iX)^{-1}X'M_iy \text{ is the BLUE.}$$

In other words, $\lambda_T\{Var(\tilde{\beta}) - Var(\hat{\beta}_{OLS})\} \geq 0$,
for some LUE $\tilde{\beta}$, and where $\lambda_T\{.\}$ denotes the smallest eigenvalue.

But then we can easily compare

$$\hat{\beta}_{FDOLS} := \left[(Ax)'Ax\right]^{-1}(Ax)'(Ay) = (x'A'Ax)^{-1}x'A'Ay,$$

with $\hat{\beta}_{OLS}$ because.

(i) $\hat{\beta}_{FD\,OLS}$ is linear since for $B := (x'A'Ax)^{-1}x'A'A$,

$$\hat{\beta}_{FD\,OLS} = By.$$

(ii) $\hat{\beta}_{FD\,OLS}$ is unbiased since

$$E\left[\hat{\beta}_{FD\,OLS}\right] = \beta + E\left[(x'A'Ax)^{-1}x'A'A\varepsilon\right] = \beta$$

and so the G/M theorem tells us that

$$\lambda_T \left\{ Var\left(\hat{\beta}_{FD\,OLS}\right) - Var\left(\hat{\beta}_{OLS}\right) \right\} \geqslant 0 ,$$

that is, $\hat{\beta}_{OLS}$ is a "better" LUE than $\hat{\beta}_{FD\,OLS}$.

APPENDIX A: COMMON QUESTION FROM STUDENTS

"RAGUIR, I DON'T UNDERSTAND WHY YOU ARE USING ALL THESE FORMULAS WITH "M_i" IN THEM...?"

Consider our model again: $y = Z\theta + \varepsilon$ where $Z := [i \ X]$ and $\theta := [\beta_0, \beta']'$.

The OLS estimator would be $\hat{\theta} := (Z'Z)^{-1}Z'y$, a $(K+1)\times 1$ vector.

- If we compare $Var(\hat{\theta})$ with $Var(\hat{\beta}_{FD\,OLS})$, that would just be silly because the former is $(K+1$ by $K+1)$ and the latter is $(K$ by $K)$. We can't even compute $Var(\hat{\theta}) - Var(\hat{\beta}_{FD\,OLS})$.
- If we compared $Var((X'X)^{-1}X'y)$ with $Var(\hat{\beta}_{FD\,OLS})$, we are indeed able to do so, but $(X'X)^{-1}X'y$ is NOT the right way to estimate β. You can't just ignore i.
- The only correct way is to run a partitioned regression.

APPENDIX B: EXTRA QUESTION TO TEST YOURSELF

IMAGINE YOU ARE THE EC402 T.A. FOR 5 MINS. YOUR STUDENT ASKS YOU:

Clearly, $\hat{\beta}_{FD\,OLS}$ & $\hat{\beta}_{OLS}$ are both OLS estimators. To see this, note: they both have the usual $(\tilde{x}'\tilde{x})^{-1}\tilde{x}'\tilde{y}$ form, where for

LEVELS : $\tilde{y} := M_i y$; $\tilde{x} := M_i x$, and for

FD : $\tilde{y} := Ay$; $\tilde{x} := Ax$.

That is, $\hat{\beta}_{OLS} = [(M_i x)' M_i x]^{-1}(M_i x)' M_i y = (x' M_i x)^{-1} x' M_i y$, and

$$\hat{\beta}_{FD\,OLS} = [(Ax)' Ax]^{-1}(Ax)' Ay = (x' A'A x)^{-1} x' A'A y .$$

The G/M theorem tells us that OLS estimators are BLUE (under suitable assumptions). However, here, we are using the G/M theorem to say that one OLS estimator is better than another? Strange, isn't it? Explain, please!

QUESTION 3

(Q3) $\underset{T\times 1}{y} = \underset{T\times K_1}{X_1}\beta_1 + \underset{T\times K_2}{X_2}\beta_2 + \varepsilon$ — ①

Assume A1: $\text{rank}([X_1, X_2]) = K_1 + K_2$,
A2: Model ① holds and $E(\varepsilon) = 0$,
A3F: $[X_1, X_2]$ non-stochastic.

(a) $\hat{\gamma}_1 := (X_1'X_1)^{-1}X_1'y \overset{\text{WHY?}}{=} \beta_1 + (X_1'X_1)^{-1}X_1'\varepsilon + (X_1'X_1)^{-1}X_1'X_2\beta_2$.

$\Rightarrow \quad E[\hat{\gamma}_1] \overset{\text{WHY?}}{=} \beta_1 + (X_1'X_1)^{-1}X_1'X_2\beta_2 \neq \beta_1$ unless $X_1'X_2 = 0$ or $\beta_2 = 0$.

(b) $\text{Var}(\hat{\gamma}_1) \overset{\text{WHY?}}{=} E\left[(\hat{\gamma}_1 - E(\hat{\gamma}_1))(\hat{\gamma}_1 - E(\hat{\gamma}_1))'\right]$

$\overset{\text{WHY?}}{=} E\left[(X_1'X_1)^{-1}X_1'\varepsilon\,\varepsilon'X_1(X_1'X_1)^{-1}\right]$

$\overset{\text{WHY?}}{=} (X_1'X_1)^{-1}X_1'E[\varepsilon\varepsilon']X_1(X_1'X_1)^{-1}$

I've just incurred a 2 mark penalty. Why? $= (X_1'X_1)^{-1}X_1'\sigma_\varepsilon^2 I_T X_1(X_1'X_1)^{-1} = \sigma_\varepsilon^2(X_1'X_1)^{-1}$

(c) $\hat{\sigma}^2_\varepsilon := (y - X_1\hat{\gamma}_1)'(y - X_1\hat{\gamma}_1) \,/\, (T - K_1)$

Step 1: $(y - X_1\hat{\gamma}_1) = y - X_1(X_1'X_1)^{-1}X_1'y = [I_T - X_1(X_1'X_1)^{-1}X_1']y = M_{X_1}y$

$= M_{X_1}[X_1\beta_1 + X_2\beta_2 + \varepsilon] = M_{X_1}X_2\beta_2 + M_{X_1}\varepsilon$.

Step 2: $(T - K_1)\hat{\sigma}^2_\varepsilon = (M_{X_1}X_2\beta_2 + M_{X_1}\varepsilon)'(M_{X_1}X_2\beta_2 + M_{X_1}\varepsilon)$

$= \beta_2'X_2'M_{X_1}X_2\beta_2 + \beta_2'X_2'M_{X_1}\varepsilon + \varepsilon'M_{X_1}X_2\beta_2 + \varepsilon'M_{X_1}\varepsilon$.

Step 3: $E(\varepsilon'M_{X_1}\varepsilon) = \dots$

$= \dots$

$= \sigma^2_\varepsilon(T - K_1)$.

This is actually a proof of separate interest on its own. I provide it below in full detail in "Appendix D".

Step 4: $E[\hat{\sigma}^2_\varepsilon] = \sigma^2_\varepsilon + \frac{(X_2\beta_2)'M_{X_1}(X_2\beta_2)}{(T - K_1)} \geq \sigma^2_\varepsilon$ because ...

STATE THE REASON PRECISELY

WHY?

In other words, $\hat{\sigma}^2_\varepsilon$ systematically over- under- estimates σ^2_ε.

choose one.

(d) If X_{1t} and X_{2t} are uncorrelated, we will have $(X_1'X_1)^{-1}X_1'X_2 = 0$

$\therefore$ $\hat{\gamma}_1$ would be unbiased for β_1.

However $M_{X_1}X_2$ would be "large" relative to the correlated case.

$\therefore$ the (upward) bias in $\hat{\sigma}^2_\varepsilon$ would be "the worst possible" relative to the correlated case.

In case you need help with the "NOTE" part of Vassilis' solution:

- Consider a correctly specified model (i.e. $\beta_2 \neq 0$ and $X_1' X_2 \neq 0$):

$$y = X_1 \beta_1 + X_2 \beta_2 + \varepsilon, \text{ with } \varepsilon \sim N(0, \sigma^2 I_T)$$

and $\operatorname{rank}([X_1, X_2]) = k_1 + k_2$ and $[X_1, X_2]$ non-stochas.

- We've seen that $E(\hat{\gamma}_1) = \beta_1 + (X_1' X_1)^{-1} X_1' X_2 \beta_2$ (Biased)

and $\operatorname{Var}(\hat{\gamma}_1) = \sigma^2 (X_1' X_1)^{-1}$.

- Further, we know that for $\hat{\beta}_{1,LSE} := (X_1' M_{X_2} X_1)^{-1} X_1' M_{X_2} y$,

$$E(\hat{\beta}_{1,LSE}) \overset{\text{WHY?}}{=} \beta_1 \qquad \text{(Unbiased)}$$

and $\operatorname{Var}(\hat{\beta}_{1,LSE}) \overset{\text{WHY?}}{=} \sigma^2 (X_1' M_{X_2} X_1)^{-1}$.

So, let's compare $X_1' M_{X_2} X_1$ with $X_1' X_1$:

In general, we'd expect $\lambda_T \left\{ (X_1' X_1) - (M_{X_2} X_1)'(M_{X_2} X_1) \right\} > 0$ if $X_1' X_2 \neq 0$,

i.e. we'd expect $\lambda_T \left\{ Var(\hat{\beta}_{1,LSE}) - Var(\hat{\gamma}_1) \right\} > 0$,

where $\lambda_T\{.\}$ is the smallest eigenvalue.

This is one reason "kitchen sink regressions" are generally not a good idea.

(FYI: this bias/variance tradeoff appears in many different guises throughout Statistics.)

Appendix C: MSE Review

- Say $\hat{\theta}$ is an estimator for θ. We want to analyse its properties; Is it even a decent estimator?

(i) - one person may say "I care about bias". Is $\hat{\theta}$ centred on the truth?
i.e. Is $|E(\hat{\theta}) - \theta|$ low or high?

(ii) - another may say "I care about variance". Is $\hat{\theta}$ clustered (or dispersed) around its centre?
i.e. Is $\text{Var}(\hat{\theta})$ low or high?

(iii) - a third may say "I care about MSE". Are squared errors in estimation of θ (using $\hat{\theta}$) zero ... on average?
i.e. Is $E[(\hat{\theta} - \theta)^2]$ low or high?

- Turns out that person (iii)'s criterion is not that different from the other two:

$$
\begin{aligned}
\text{MSE}(\hat{\theta}) := E[(\hat{\theta} - \theta)^2] &= E[\hat{\theta}^2 - 2\theta\hat{\theta} + \theta^2] = E(\hat{\theta}^2) - 2\theta E(\hat{\theta}) + \theta^2 \\
&= \text{Var}(\hat{\theta}) + (E(\hat{\theta}))^2 - 2\theta E(\hat{\theta}) + \theta^2 \\
&= \text{Var}(\hat{\theta}) + [E(\hat{\theta}) - \theta]^2 \\
&= \text{Var}(\hat{\theta}) + \text{Bias}^2(\hat{\theta}).
\end{aligned}
$$

(Now, have a look at the last line of Vassilis' solution.)

Appendix D

$$E[\varepsilon' M_X \varepsilon] \overset{\text{why?}}{=} E[\mathrm{Tr}\{\varepsilon' M_X \varepsilon\}] \overset{\text{why?}}{=} E[\mathrm{Tr}\{\varepsilon\varepsilon' M_X\}] \overset{\text{why?}}{=} \mathrm{Tr}\{E[\varepsilon\varepsilon'] M_X\}$$

$$\overset{\text{why?}}{=} \mathrm{Tr}\{\sigma_\varepsilon^2 I_T M_X\} = \sigma_\varepsilon^2 \mathrm{Tr}\{M_X\} = \sigma_\varepsilon^2 \mathrm{Tr}\{I_T - X(X'X)^{-1}X'\}$$

$$= \sigma_\varepsilon^2 [\mathrm{Tr}\{I_T\} - \mathrm{Tr}\{X(X'X)^{-1}X'\}] = \sigma_\varepsilon^2 [\mathrm{Tr}\{I_T\} - \mathrm{Tr}\{(X'X)(X'X)^{-1}\}]$$

$$\overset{\text{why?}}{=} \sigma_\varepsilon^2 [\mathrm{Tr}\{I_T\} - \mathrm{Tr}\{I_K\}] = \sigma_\varepsilon^2 (T-K).$$

State the two red "why?"s explicitly in your answer!

For us, $X := X_1$ and $K = K_1$ $\therefore$ $E[\varepsilon' M_{X_1} \varepsilon] = \sigma_\varepsilon^2 (T - K_1)$ as required.

These equations are useful for instance at the end of Vassilis' solution to Q3 of PS2.

QUESTION 4

(Q4) $y = X\beta + \varepsilon$, $\varepsilon | X \sim (0, \sigma^2 \Omega)$

with $\text{rank}(X'X) = k$, and y is $T \times 1$.

For $t, s = 1, \ldots, T$, consider the cases below:

(a) $$E[\varepsilon_t \varepsilon_s | X] = \begin{cases} \theta_0 + \theta_1 x_{3t}^2 + \theta_2 x_{5t}^{-4}, & t = s \\ 0, & t \neq s. \end{cases}$$

(b) $$E[\varepsilon_t \varepsilon_s | X] = \begin{cases} \sigma_v^2 \eta_t, & t = s \\ 0, & t \neq s, \end{cases}$$

where σ_v^2 is unknown but η_t is known.

(a)(i) let $E[\varepsilon\varepsilon' | X] := c^2\Omega$,

where $c^2 := 1$, and

$\Omega \equiv \Omega(\theta) := \operatorname{diag}\{z_1'\theta, \dots, z_T'\theta\}$ for $\theta := (\theta_0, \theta_1, \theta_2)'$ and

$z_t := (1, x_{3t}^2, x_{5t}^{\cdot 4})'$ for $t = 1, \dots, T$.

(ii) Step 1. Define $\hat{\varepsilon} := y - X(X'X)^{-1}X'y$,

and for $t = 1, \dots, T$, let $\hat{\varepsilon}_t$ denote the t^{th} element of $\hat{\varepsilon}$.

Step 2. Define $\hat{\theta} := \left(\sum_{t=1}^{T} z_t z_t'\right)^{-1} \sum_{t=1}^{T} z_t \hat{\varepsilon}_t^2$

Step 3. Define $\hat{\Omega} := \Omega(\hat{\theta}) = \operatorname{diag}\{z_1'\hat{\theta}, \dots, z_T'\hat{\theta}\}$

Step 4. Define $\hat{\beta}_{FGLS} := (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}y$.

(b) (i) let $E[\varepsilon\varepsilon'|X] := c^2\Omega$,

where $c^2 := \sigma_v^2$, and

$$\Omega = \Omega(\theta) := \operatorname{diag}\{\eta_1, \dots, \eta_T\} \text{ for } \theta := (\eta_1, \dots, \eta_T)'.$$

(ii) Define $\hat{\beta}_{FGLS} := (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y \equiv \hat{\beta}_{GLS}$ since θ is known.

For $t \in \mathbb{Z}$, consider the case below:

(c) $\varepsilon_t = \rho \varepsilon_{t-1} + v_t$, $|\rho| < 1$, $v_t \sim \text{IID}$, $E[v_t | X] = 0$, $Var[v_t | X] = \sigma_v^2 < \infty$.

(i) For this (cov. stationary) process, since

$$E[\varepsilon_t \varepsilon_{t-h} | X] = \begin{cases} \sigma_\varepsilon^2, & \text{if } h = 0 \text{ for some } 0 < \sigma_\varepsilon^2 < \infty \\ E\left[\left(\rho^h \varepsilon_{t-h} + \sum_{k=0}^{h-1} \rho^k v_{t-k}\right) \varepsilon_{t-h} \middle| X\right], & h > 0 \\ E\left[\varepsilon_t \left(\rho^{-h} \varepsilon_t + \sum_{k=0}^{-h-1} \rho^k v_{t+k}\right) \middle| X\right], & h < 0 \end{cases}$$

and $E[\varepsilon_m v_n | X] = 0$ for any integers $m < n$, it follows that

that $E[\varepsilon_t \varepsilon_{t-h} | X] = \rho^{|h|} \sigma_\varepsilon^2$ for $h = -\infty, \ldots, 0, \ldots, \infty$.

Thus, for $t = 1, \ldots, T$, we let

$T \times T$ matrix $E\left[\varepsilon\varepsilon' \mid X\right] := c^2 \Omega$,

where $c^2 := \sigma_\varepsilon^2$

$$\Omega = \Omega(\rho) := \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{T-1} \\ \rho & 1 & \rho & \cdots & \rho^{T-2} \\ \rho^2 & \rho & 1 & & \rho^{T-3} \\ \vdots & \vdots & \vdots & \ddots & \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & & 1 \end{bmatrix}$$

[Note: People who have studied time series will know that $\sigma_\varepsilon^2 = \sigma_v^2 / (1-\rho^2)$.

Test yourself: why did I not bother to torture the non-time-series students by requiring them to compute σ_ε^2 for this question?]

(ii) Step 1. Define $\hat{\varepsilon} := y - X(X'X)^{-1}X'y$,

and for $t = 1, \dots, T$, let $\hat{\varepsilon}_t$ denote the t^{th} element of $\hat{\varepsilon}$.

Step 2. Define $\hat{\rho} := \left(\sum_{t=2}^{T} \hat{\varepsilon}_{t-1} \hat{\varepsilon}_{t-1} \right)^{-1} \sum_{t=2}^{T} \hat{\varepsilon}_{t-1} \hat{\varepsilon}_{t}$.

Step 3. Define $\hat{\beta}_{FGLS} := (X'\hat{\Omega}^{-1}X)^{-1} X'\hat{\Omega}^{-1}y$

where $\hat{\Omega} := \Omega(\hat{\rho})$ as defined above.

For $t \in \mathbb{Z}$, consider the case below:

(d) $\varepsilon_t = v_t + \lambda v_{t-1}$, $v_t \sim IID$, $E[v_t|X] = 0$, $Var[v_t|X] = \sigma_v^2 < \infty$

(i) For this (Cov. stationary) process, since

$$E[\varepsilon_t \varepsilon_{t-h}|X] = E[(v_t + \lambda v_{t-1})(v_{t-h} + \lambda v_{t-h-1})|X], \; h \in \mathbb{Z}$$

and $E[v_m v_n|X] = 0$ for any integers $m \neq n$, it follows that

$$E[\varepsilon_t \varepsilon_{t-h}|X] = \begin{cases} \sigma_v^2(1+\lambda^2) & , \text{if } h = 0 \\ \lambda \sigma_v^2 & , |h| = 1 \\ 0 & , |h| > 1 \end{cases},$$

for $h = -\infty, \dots, 0, \dots, \infty$.

Thus, the ACF of ε_t is given by

$$\rho_\varepsilon(h) := \begin{cases} 1 & , \ h = 0 \\ \lambda / (1+\lambda^2) & , \ |h| = 1 \\ 0 & , \ \text{otherwise} \end{cases},$$

for $h \in \mathbb{Z}$

Thus, for $t=1,\dots,T$, we can let

$T\times T$ matrix $E[\varepsilon\varepsilon' | X] := c^2\Omega$, where

$c^2 := (1+\lambda^2)\sigma_v^2$ and

$$\Omega \equiv \Omega(\rho) := \begin{bmatrix} 1 & \rho & 0 & \cdots & 0 \\ \rho & 1 & \rho & \cdots & 0 \\ 0 & \rho & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

where $\rho := \rho_\varepsilon(-1) = \rho_\varepsilon(1) = \dfrac{\lambda}{1+\lambda^2}$.

(ii) Step 1. Define $\hat{\varepsilon} := y - X(X'X)^{-1}X'y$ and let $\hat{\varepsilon}_t$ denote the t^{th} element of $\hat{\varepsilon}$, for $t = 1, \dots, T$.

Step 2. Define $\hat{\rho} := \sum_{t=2}^{T} \hat{\varepsilon}_{t-1}\hat{\varepsilon}_t \Big/ \sum_{t=1}^{T} \hat{\varepsilon}_t^2$

Step 3. Define $\hat{\beta}_{FGLS} := (X'\hat{\Omega}^{-1}X)^{-1} X'\hat{\Omega}^{-1}y$

where $\hat{\Omega} := \Omega(\hat{\rho})$ as defined above.