Exo8 corrigé v2 A2011

Q1. Nous avons le modèle suivant: $y_i = \beta_1 + x_{i2}\beta_2 + x_{i3}\beta_3 + \varepsilon_i$ où les aléas sont distribués tel que

$$\varepsilon \sim N\left(0_{n\times 1}\ ,\ \sigma^2\begin{bmatrix}\Lambda_1 & 0\\ 0 & \Lambda_2\end{bmatrix}_{n\times n}\right)$$

avec les matrices diagonales suivantes $\Lambda_1 = c_1 I_{n_1}$ et $\Lambda_2 = c_2 I_{n_2}$, de dimension n_1 et n_2 respectivement, et où c_1 et c_2 sont des constantes.

a) Trouvez une manière efficace d'estimer β sans biais.
Dérivez l'estimateur à partir de son critère (en faisant les conditions de premier ordre, FOC).
De plus, écrivez sa variance détaillée (la matrice de variance-covariance de votre estimateur $\hat{\beta}$).
Discutez des propriétés de cet estimateur dans ce contexte.

Ici on peut faire GLS ou ML

Avec $X = \begin{bmatrix}\iota & x_2 & x_3\end{bmatrix}$ et $\beta = \begin{bmatrix}\beta_1\\ \beta_2\\ \beta_3\end{bmatrix}$

Avec $\Sigma = \sigma^2\Omega = \sigma^2\begin{bmatrix}\Lambda_1 & 0\\ 0 & \Lambda_2\end{bmatrix}_{n\times n} = \begin{bmatrix}c_1 I_{n_1} & 0\\ 0 & c_2 I_{n_2}\end{bmatrix}$ qui est connu

L'estimateur GLS (MCG) est donné par

$$\hat{\beta}_{GLS} = \underset{\beta\in\mathbb{R}^K}{\arg\min}\, S(\beta) = (y - X\beta)'\Omega^{-1}(y - X\beta)$$
$$\text{ou bien par}$$
$$\hat{\beta}_{GLS} = \underset{\beta\in\mathbb{R}^K}{\arg\min}\, S(\beta) = (y - X\beta)'\Sigma^{-1}(y - X\beta)$$

Le critère se réécrit comme

$$\begin{aligned}s(\beta) &= (y - X\beta)'\Omega^{-1}(y - X\beta) = (y - X\beta)'(\Omega^{-1}y - \Omega^{-1}X\beta)\\ &= y'\Omega^{-1}y - \beta'X'\Omega^{-1}y - y'\Omega^{-1}X\beta + \beta'X'\Omega^{-1}X\beta\\ &= y'\Omega^{-1}y - 2y'\Omega^{-1}X\beta + \beta'X'\Omega^{-1}X\beta\end{aligned}$$

The FOC are given by:

$\dfrac{\partial s(\beta)}{\partial\beta} = (-2X'\Omega^{-1}y + 2X'\Omega^{-1}X\ \hat{\beta}_{GLs}) = 0$ which gives us

$$\hat{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y = (X'\tilde{P}'\underbrace{\tilde{P}X}_{\tilde{X}})^{-1}X'\tilde{P}'\underbrace{\tilde{P}y}_{\tilde{y}} = \underbrace{(\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{y}}_{\substack{\text{MCO avec le modèle} \\ \text{transformé par } \tilde{P}}}$$

$$= (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y = (X'P'\underbrace{PX}_{X^*})^{-1}X'P'\underbrace{Py}_{y^*} = \underbrace{(X^{*\prime}X^*)^{-1}X^{*\prime}y^*}_{\substack{\text{MCO avec le modèle} \\ \text{transformé par } P}}$$

$$\Omega = \begin{bmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{bmatrix} = \begin{bmatrix} c_1 I_{n_1} & 0 \\ 0 & c_2 I_{n_2} \end{bmatrix}$$

$$\Omega^{-1} = \begin{bmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{bmatrix}^{-1} = \begin{bmatrix} c_1 I_{n_1} & 0 \\ 0 & c_2 I_{n_2} \end{bmatrix}^{-1} = \begin{bmatrix} \frac{1}{c_1} I_{n_1} & 0 \\ 0 & \frac{1}{c_2} I_{n_2} \end{bmatrix}$$

Où $\tilde{P} = \Omega^{-1/2} = \begin{bmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{bmatrix}^{-1/2} = \begin{bmatrix} c_1 I_{n_1} & 0 \\ 0 & c_2 I_{n_2} \end{bmatrix}^{-1/2} = \begin{bmatrix} \frac{1}{\sqrt{c_1}} I_{n_1} & 0 \\ 0 & \frac{1}{\sqrt{c_2}} I_{n_2} \end{bmatrix}$

On peut écrire $X = \begin{bmatrix} X^{[n_1]} \\ X^{[n_2]} \end{bmatrix}$ et $y = \begin{bmatrix} y^{[n_1]} \\ y^{[n_2]} \end{bmatrix}$

Ainsi on estimera le modèle transformé par MCO avec la matrice P où $P'P = \Sigma^{-1} = (\sigma^2\Omega)^{-1}$

$$Py = PX\beta + P\varepsilon,$$

Que l'on réécrit comme

$$y^* = X^*\beta + \varepsilon^*.$$

ou bien avec la matrice $\tilde{P}$ où $\tilde{P}'\tilde{P} = \Omega^{-1}$

$$\tilde{P}y = \tilde{P}X\beta + \tilde{P}\varepsilon,$$

Que l'on réécrit comme

$$\tilde{y} = \tilde{X}\beta + \tilde{\varepsilon}.$$

La variance est données par

$$\operatorname{var}(\hat{\beta}_{GLS}) = \sigma^2(X'\Omega^{-1}X)^{-1} = \sigma^2(X'\tilde{P}'\tilde{P}X)^{-1} = \sigma^2(\tilde{X}'\tilde{X})^{-1} = (X'\Sigma^{-1}X)^{-1} = (X'P'PX)^{-1} = (X^{*\prime}X^*)^{-1}$$

$$\operatorname{var}(\hat{\beta}_{GLS}) = \sigma^2 (X'\Omega^{-1}X)^{-1} = \sigma^2 \left(\begin{bmatrix} X^{[n_1]} \\ X^{[n_2]} \end{bmatrix}' \begin{bmatrix} \frac{1}{c_1} I_{n_1} & 0 \\ 0 & \frac{1}{c_2} I_{n_2} \end{bmatrix} \begin{bmatrix} X^{[n_1]} \\ X^{[n_2]} \end{bmatrix} \right)^{-1}$$

$$= \sigma^2 \left(\begin{bmatrix} X^{[n_1]} \\ X^{[n_2]} \end{bmatrix}' \begin{bmatrix} \frac{1}{c_1} X^{[n_1]} + 0 \\ 0 + \frac{1}{c_2} X^{[n_2]} \end{bmatrix} \right)^{-1} = \sigma^2 \left(\begin{bmatrix} X^{[n_1]}{}' & X^{[n_2]}{}' \end{bmatrix} \begin{bmatrix} \frac{1}{c_1} X^{[n_1]} \\ \frac{1}{c_2} X^{[n_2]} \end{bmatrix} \right)^{-1} = \sigma^2 \left[X^{[n_1]}{}' \frac{1}{c_1} X^{[n_1]} + X^{[n_2]}{}' \frac{1}{c_2} X^{[n_2]} \right]^{-1}$$

$E(\hat{\beta}_{GLS}) = \beta_0$ non-biaisé et consistant
Variable minimale (car MCO avec le modèle transformé ou GLS avec la vraie variance est BLUE)
Estimateur efficace.

Dans ce cas $\hat{\beta}_{GLS} \sim N\left(\beta \ , \sigma^2 (X'\Omega^{-1}X)^{-1}\right)$ car on a la normalité des aléas

On pourrait aussi utiliser MLE.

So we get that $f(y) = f(\varepsilon) \left| \frac{\partial \varepsilon}{\partial y'} \right| = f(\varepsilon)$

If we had $\varepsilon \sim N(0, \sigma^2 \Omega)$, the density of $\varepsilon_i \sim N(0, \sigma^2)$ would be given by the PDF of the normal distribution:

In matrix notation we thus have: $f(\varepsilon) = \frac{1}{(2\pi)^{n/2}} \left| \Sigma \right|^{-1/2} \exp\left(\frac{-1}{2} \frac{\varepsilon' \Sigma^{-1} \varepsilon}{1} \right) = \frac{1}{(2\pi)^{n/2}} \left| \sigma^2 \Omega \right|^{-1/2} \exp\left(\frac{-1}{2} \frac{\varepsilon' \Omega^{-1} \varepsilon}{\sigma^2} \right)$

Playing with danger, here we can write the likelihood function for y as:

$L(\theta \mid y , X) \equiv f(y) = \frac{1}{(2\pi\sigma^2)^{n/2}} \left| \Omega \right|^{-1/2} \exp\left(\frac{-1}{2} \frac{(y - X\beta)'\Omega^{-1}(y - X\beta)}{\sigma^2} \right)$ where $\theta = \begin{bmatrix} \beta \\ \sigma^2 \end{bmatrix}$

$$\max_{\theta \in \Theta} \ln L(\theta \mid y , X) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln\left| \sigma^2 \right| - \frac{1}{2} \ln\left| \Omega \right| - \frac{1}{2\sigma^2} (y - X\beta)'\Omega^{-1}(y - X\beta)$$

Since $\left| \sigma^2 \Omega \right| = (\sigma^2)^n \left| \Omega \right|$

Note that the term in the brackets in the last term is

$$(y - X\beta)'\Omega^{-1}(y - X\beta) = (y - X\beta)'(\Omega^{-1}y - \Omega^{-1}X\beta)$$
$$= y'\Omega^{-1}y - \beta' X'\Omega^{-1}y - y'\Omega^{-1}X\beta + \beta' X'\Omega^{-1}X\beta$$
$$= y'\Omega^{-1}y - 2y'\Omega^{-1}X\beta + \beta' X'\Omega^{-1}X\beta$$

Ici si σ^2 est connu, on aurait pas besoin de l'estimer.

The FOC are given by:

$$\frac{\partial \ln L(\beta \mid y, X)}{\partial \beta} = -\frac{1}{2\sigma^2}(-2X'\Omega^{-1}y + 2X'\Omega^{-1}X\hat{\beta}_{ML}) = 0 = \frac{1}{\sigma^2}(X'\Omega^{-1}y + X'\Omega^{-1}X\hat{\beta}_{ML}) = 0$$ which gives us

$$\hat{\beta}_{ML} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y = (X'\tilde{P}'\underbrace{\tilde{P}X}_{\tilde{X}})^{-1}X'\tilde{P}'\underbrace{\tilde{P}y}_{\tilde{y}} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{y}$$
$$= (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y = (X'P'\underbrace{PX}_{X^*})^{-1}X'P'\underbrace{Py}_{y^*} = (X^*{}'X^*)^{-1}X^*{}'y^*$$

Les propriétés seraient les mêmes. $E(\hat{\beta}_{ML}) = \beta_0$ non-biaisé et consistant. Variable minimale (car MCO avec le modèle transformé ou GLS avec la vraie variance est BLUE) Estimateur efficace.

b) Construisez un test de restrictions linéaires de la nulle dans le cas où σ inconnu et dans le cas où σ est connu

$H_0 : \begin{bmatrix}\beta_2 \\ \beta_3\end{bmatrix} = \begin{bmatrix}1 \\ 1\end{bmatrix}$ vs $H_1 : \begin{bmatrix}\beta_2 \\ \beta_3\end{bmatrix} \neq \begin{bmatrix}1 \\ 1\end{bmatrix}$

N'oubliez pas de bien indiquer la distribution que suit ce test et les règles de décision.

Test de Fisher avec l'estimé $\hat{\sigma}^2$ si σ est inconnu

$H_0 : R\beta = \begin{bmatrix}0 & 1 & 0 \\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}\beta_1 \\ \beta_2 \\ \beta_3\end{bmatrix} = \begin{bmatrix}1 \\ 1\end{bmatrix} = r$ **vs** $H_1 : R\beta = \begin{bmatrix}0 & 1 & 0 \\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}\beta_1 \\ \beta_2 \\ \beta_3\end{bmatrix} = \begin{bmatrix}1 \\ 1\end{bmatrix} \neq r$

Avec des aléas normaux et un σ inconnu on peut construire le test de Fisher avec la formule suivante à partir de l'estimateur GLS, ici il faut tenir compte du fait que $\Sigma = \sigma^2\Omega = \sigma^2\begin{bmatrix}\Lambda_1 & 0 \\ 0 & \Lambda_2\end{bmatrix}$:

La matrice de restrictions linéaire est définie comme $R = \begin{bmatrix}0 & 1 & 0 \\ 0 & 0 & 1\end{bmatrix}$ et $J = rang(R) = 2$

$$F_{GLS} = \frac{\left(R\hat{\beta}_{GLS} - r\right)'\left(R\left(X'\Omega^{-1}X\right)^{-1}R'\right)^{-1}\left(R\hat{\beta}_{GLS} - r\right)/J}{\hat{\sigma}^2_{GLS}} \overset{H_0}{\sim} F(J, n-K).$$

avec $\hat{\sigma}^2_{GLS} = \hat{\varepsilon}'\Omega^{-1}\hat{\varepsilon} / (n-K)$

On pourrait aussi construire le test à partir d'un modèle transformé $\tilde{P}y = \tilde{P}X\beta + \tilde{P}\varepsilon$

$$\Rightarrow \hat{\beta}_{OLS-\tilde{P}} = ((\tilde{P}X)'(\tilde{P}X))^{-1}(\tilde{P}X)'\tilde{P}y = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{y} = \hat{\beta}_{GLS}$$

$$F_{OLS-\tilde{P}} = \frac{\left(R\hat{\beta}_{OLS-\tilde{P}} - r\right)'\left(R\left((\tilde{P}X)'(\tilde{P}X)\right)^{-1}R'\right)^{-1}\left(R\hat{\beta}_{OLS-\tilde{P}} - r\right)/J}{\hat{\sigma}^2_{OLS-\tilde{P}}} \overset{H_0}{\sim} F(J, n-K).$$

Avec $\hat{\sigma}^2_{OLS-\tilde{P}} = \hat{\tilde{\varepsilon}}_{OLS-\tilde{P}}{}'\hat{\tilde{\varepsilon}}_{OLS-\tilde{P}} / (n-K) = (\tilde{P}y - \tilde{P}X\hat{\beta}_{OLS-\tilde{P}})'(\tilde{P}y - \tilde{P}X\hat{\beta}_{OLS-\tilde{P}}) / (n-K)$

On pourrait aussi prendre le modèle transformé par la matrice P pour construire le test.

Règle de décision (niveau de signification de 0.05)

On ne rejette pas H_0 si $F \leq F_\alpha(J, n-K)$

On rejette H_0 en faveur de H_1 si $F > F_\alpha(J, n-K)$

Si par contre σ^2 est connu. On aura une distribution $\chi^2_{(2)}$. Ce sera un test de Wald.

$$Wald = \left(R\hat{\beta}_{GLS} - r\right)'\left(R\sigma^2\left(X'\Omega^{-1}X\right)^{-1}R'\right)^{-1}\left(R\hat{\beta}_{GLS} - r\right) \overset{H_0}{\sim} \chi^2_{(2)}$$

On peut aussi construire le test à partir de l'estimation du modèle transformé estimé par OLS, le résultat sera identique :

$$Wald = \left(R\hat{\beta}_{OLS-\tilde{P}} - r\right)'\left(R\sigma^2\left(\tilde{X}'\tilde{X}\right)^{-1}R'\right)^{-1}\left(R\hat{\beta}_{OLS-\tilde{P}} - r\right) \overset{H_0}{\sim} \chi^2_{(2)}$$

Règle de décision (niveau de signification de 0.05)

On ne rejette pas H_0 si $Wald \leq \chi^2_{(2)\alpha}$

On rejette H_0 en faveur de H_1 si $Wald > \chi^2_{(2)\alpha}$

c) Comme dans la pratique vous n'êtes jamais certain de la normalité, vous faites un test de Jarques-Bera et vous obtenez une valeur de 1001 pour le ratio $\left(\widehat{\mu}_3 \ / \ \hat{\sigma}^3\right)$ obtenu à partir des résidus de votre régression. Quelles sont les implications pour vos réponses en 3a) et 3b)

Il faut utiliser les tables asymptotiques des tests car on perd la normalité de $\hat{\beta}_{GLS}$ et de $\hat{\beta}_{ML}$ en échantillon fini.
Toujours non-biaisé et consistant. Toujours variance minimale donc BLUE.
Estimateur efficace.

Q4-

$R(\theta) = \beta_1 \cdot \beta_2 = 2 = r$

The statistic of the test would be given by

$$W = \left(R(\hat{\beta}) - r\right)'\left[\widehat{var}\left(R(\hat{\beta})\right)\right]^{-1}\left(R(\hat{\beta}) - r\right) \xrightarrow{d} \chi^2(J)$$

To obtain the variance of a non-linear function we use the Delta method, which takes a taylor expansion such that

$$R(\hat{\beta}) = R(\beta) + \left(\frac{\partial R(\beta)}{\partial \beta'}\right)_{\beta=\hat{\beta}} (\hat{\beta} - \beta) + ...$$

where the subscript $\beta = \hat{\beta}$ means that the gradient is evaluated at $\beta = \hat{\beta}$

From there

$$\text{var}\left(R(\hat{\beta})\right) = \text{var}\left(R(\hat{\beta}) - R(\beta)\right) = \text{var}\left(\left(\frac{\partial R(\hat{\beta})}{\partial \beta'}\right)_{\beta=\hat{\beta}} (\hat{\beta} - \beta)\right) = E\left[\underbrace{\left(\frac{\partial R(\hat{\beta})}{\partial \beta'}\right)_{\beta=\hat{\beta}}}_{J\times K} \underbrace{(\hat{\beta} - \beta)(\hat{\beta} - \beta)'}_{K\times K} \underbrace{\left(\frac{\partial R(\hat{\beta})}{\partial \beta'}\right)'_{\beta=\hat{\beta}}}_{K\times J}\right.$$

$$= \left(\frac{\partial R(\hat{\beta})}{\partial \beta'}\right)_{\beta=\hat{\beta}} \text{var}(\hat{\beta}) \left(\frac{\partial R(\hat{\beta})}{\partial \beta'}\right)'_{\beta=\hat{\beta}}$$

Which we use to built the Wald test.

In our example $\left(\frac{\partial R(\hat{\beta})}{\partial \beta'}\right) = \left[\frac{\partial R(\hat{\beta})}{\partial \beta_1} \quad \frac{\partial R(\hat{\beta})}{\partial \beta_2}\right]_{\beta=\hat{\beta}} = \left[\hat{\beta}_2 \quad \hat{\beta}_1\right]$

And $\widehat{var}(\hat{\beta}) = \begin{bmatrix} \widehat{var}(\hat{\beta}_1) & \widehat{\text{cov}}(\hat{\beta}_1, \hat{\beta}_2) \\ \widehat{\text{cov}}(\hat{\beta}_2, \hat{\beta}_1) & \widehat{var}(\hat{\beta}_2) \end{bmatrix}$

So we would get

$$W = \left(\hat{\beta}_1 \cdot \hat{\beta}_2 - 2\right)'\left[\left[\hat{\beta}_2 \quad \hat{\beta}_1\right]\begin{bmatrix} \widehat{var}(\hat{\beta}_1) & \widehat{\text{cov}}(\hat{\beta}_1, \hat{\beta}_2) \\ \widehat{\text{cov}}(\hat{\beta}_2, \hat{\beta}_1) & \widehat{var}(\hat{\beta}_2) \end{bmatrix}\left[\hat{\beta}_2 \quad \hat{\beta}_1\right]'\right]^{-1}\left(\hat{\beta}_1 \cdot \hat{\beta}_2 - 2\right) \xrightarrow{d} \chi^2(J)$$

With J=1 as we only have one restriction in this case.

Q5 a)

$$E\hat{\theta} = \frac{1}{g}Ex_1 + \frac{(g-1)}{g}\frac{1}{(n-1)}\sum_{i=2}^{n} Ex_i$$
$$= \frac{1}{g}\mu_X + \frac{(g-1)}{g}\frac{1}{(n-1)}\sum_{i=2}^{n}\mu_X$$
$$= \frac{1}{g}\mu_X + \frac{(g-1)}{g}\frac{1}{(n-1)}(n-1)\mu_X$$
$$= \frac{1}{g}\mu_X + \frac{(g-1)}{g}\mu_X = \mu_X$$

L'estimateur est non-biaisé

b) Calculez sa variance

$$\mathrm{var}(\hat{\theta}) = \mathrm{var}\left(\frac{1}{g}x_1 + \frac{(g-1)}{g}\frac{1}{(n-1)}\sum_{i=2}^{n} x_i\right)$$
$$= \mathrm{var}\left(\frac{1}{g}x_1\right) + \mathrm{var}\left(\frac{(g-1)}{g}\frac{1}{(n-1)}\sum_{i=2}^{n} x_i\right) \quad car\ \ i.i.d$$
$$= \left(\frac{1}{g}\right)^2 \mathrm{var}(x_1) + \left(\frac{(g-1)}{g}\frac{1}{(n-1)}\right)^2 \sum_{i=2}^{n} \mathrm{var}(x_i)$$
$$= \left(\frac{1}{g}\right)^2 \sigma_X^2 + \left(\frac{(g-1)}{g}\frac{1}{(n-1)}\right)^2 (n-1)\sigma_X^2$$
$$= \left(\left(\frac{1}{g}\right)^2 + \left(\frac{(g-1)}{g}\frac{1}{(n-1)}\right)^2 (n-1)\right)\sigma_X^2$$
$$= \left(\left(\frac{1}{g}\right)^2 + \left(\frac{(g-1)}{g}\right)^2 \frac{1}{(n-1)}\right)\sigma_X^2$$

c) Dites si l'estimateur est consistant (convergent en probabilité) en général.

$$\lim_{n\to\infty}\left(\mathrm{var}(\hat{\theta})\right) = \lim_{n\to\infty}\left(\left(\frac{1}{g}\right)^2 + \left(\frac{(g-1)}{g}\right)^2 \frac{1}{(n-1)}\right)\sigma_X^2 = \left(\frac{1}{g}\right)^2 \sigma_X^2 \neq 0$$

En général comme la variance ne tend pas vers 0 lorsque n tend vers l'infini, l'estimateur n'est pas convergent.

Par ailleurs, si on choisit $g = n$

$$\mathrm{var}(\hat{\theta}) = \mathrm{var}\left(\frac{1}{n}x_1 + \frac{(n-1)}{n}\frac{1}{(n-1)}\sum_{i=2}^{n} x_i\right) = \mathrm{var}\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \left(\frac{1}{n}\right)^2 n\sigma_X^2 = \left(\frac{1}{n}\right)\sigma_X^2$$

Alors $\lim_{n\to\infty}\left(\mathrm{var}(\hat{\theta})\right) = \lim_{n\to\infty}\left(\frac{1}{n}\right)\sigma_X^2 = 0$. Ainsi, comme l'espérance est une constante et que la variance tend vers 0.

On a la convergence en moyenne quadratique, qui implique la convergence en probabilité, ce qui nous donne un estimateur consistant lorsque $g = n$. Dans ce cas particulier, $\hat{\theta} \xrightarrow{p} \mu_X$.