Chapitre 2 : Propriétés en petits échantillons de l'estimateur des MCO

33. Soit deux variables aléatoires x et y avec :

$$E\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix} \quad et \quad V\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix} = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix}$$

a) Montrez que ρ est la corrélation entre x et y ?
b) Quelle est l'espérance et la variance de $3x$?
c) Quelle est l'espérance et la variance de $x - y$?
d) Quelle est la covariance et la corrélation entre $(x + y)$ et $(x - y)$?

34. Précisez la notion d'exogénéité stricte (forte) ? Soit un modèle :

$$\begin{cases} y_t = \alpha + \beta x_t + \varepsilon_t \\ x_t = \delta + \gamma y_{t-1} + \nu_t \end{cases} \quad t = 1,2, \dots, T.$$

avec ε et ν indépendant. Est-ce que x_t est strictement exogène dans la première équation ? Quelles sont les conséquences sur les propriétés de l'estimateur des MCO ?

35. Pourquoi utilise-t-on $N - K$ pour le dénominateur de l'estimateur des MCO de la variance : $\widehat{\sigma^2} = SCR/(N - K)$? Montrez que cet estimateur est sans biais ? Quelle est la distribution de cet estimateur sous l'hypothèse de normalité des erreurs ?

36. Démontrez que la trace de la matrice idempotente M est égale à $N - K$.

37. Soit un estimateur $\widetilde{\boldsymbol{\beta}}$ d'un vecteur de paramètre $\boldsymbol{\beta}$,

a) Prouvez que : $Var(\widetilde{\boldsymbol{\beta}}) = E[Var(\widetilde{\boldsymbol{\beta}}|\boldsymbol{X})] + Var[E(\widetilde{\boldsymbol{\beta}}|\boldsymbol{X})]$.

b) Prouvez que $Var(\widetilde{\boldsymbol{\beta}}) \geq Var(\widehat{\boldsymbol{\beta}})$.

Indices : Par définition,

$$Var(\widetilde{\boldsymbol{\beta}}|\boldsymbol{X}) = E\left[\left(\widetilde{\boldsymbol{\beta}} - E(\widetilde{\boldsymbol{\beta}}|\boldsymbol{X})\right)\left(\widetilde{\boldsymbol{\beta}} - E(\widetilde{\boldsymbol{\beta}}|\boldsymbol{X})\right)'\middle|\boldsymbol{X}\right]$$

$$Var[E(\widetilde{\boldsymbol{\beta}}|\boldsymbol{X})] = E\left\{\left(E(\widetilde{\boldsymbol{\beta}}|\boldsymbol{X}) - E(\widetilde{\boldsymbol{\beta}})\right)\left(E(\widetilde{\boldsymbol{\beta}}|\boldsymbol{X}) - E(\widetilde{\boldsymbol{\beta}})\right)'\right\}$$

Si $Var(\widetilde{\boldsymbol{\beta}}|\boldsymbol{X}) \geq Var(\widehat{\boldsymbol{\beta}}|\boldsymbol{X})$, alors $E[Var(\widetilde{\boldsymbol{\beta}}|\boldsymbol{X})] \geq E[Var(\widehat{\boldsymbol{\beta}}|\boldsymbol{X})]$

[Hayashi (2000), exercice 4, Section I.3]

38. Prouvez la propriété II.1.i : $Cov(\boldsymbol{e}, \widehat{\boldsymbol{\beta}}|\boldsymbol{X}) = \boldsymbol{0}$
[Hayashi (2000), exercice 6, Section I.3]

39. Démontrez que la variance d'un paramètre estimé par MCO peut s'écrire :

$$V(\widehat{\beta_k}|X) = \frac{\sigma^2}{SCT_k\left(1 - R_{(k)}^2\right)}$$

avec $SCT_k = \sum_{i=1}^{N}(x_{k,i} - \overline{x_k})^2$ la somme des carrés totaux de la $k^{ème}$ variable explicative, et $R_{(k)}^2$: le R^2 de la régression de x_k sur toutes les autres variables explicatives.

40. Supposons que le vrai modèle de régression multiple soit le suivant :

$$y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \varepsilon$$

Les hypothèses H1 à H5 sont satisfaites. Cependant on estime le modèle qui omet la variable x_4. Soient $\widetilde{\beta_1}, \widetilde{\beta_2}$, et $\widetilde{\beta_3}$ les estimateurs MCO de cette régression de y sur x_2 et x_3. Montrez que l'espérance de $\widetilde{\beta_2}$ (étant donné les valeurs des variables explicatives dans l'échantillon est :

$$E(\widetilde{\beta}_2) = \beta_2 + \beta_4 \frac{\sum_{i=1}^{N} u_{i2} x_{i4}}{\sum_{i=1}^{N} u_{i2}^2}$$

avec $u_{2,i}$ les résidus de la régression par MCO de x_2 sur x_3 ?
[Wooldridge (2009), exercice 3.10]

41. Considérez le modèle de régression simple : $y = \beta_1 + \beta_2 x + \varepsilon$. Sous les hypothèses de Gauss-Markov, les estimateurs traditionnels des MCO $\widehat{\beta_1}$ et $\widehat{\beta_2}$ sont des estimateurs sans biais des paramètres de la population. Soit $\widetilde{\beta_2}$ l'estimateur de lorsqu'il n'y pas de constante dans la régression.

   a) Exprimer $E\left(\widetilde{\beta_2}\right)$ en fonction des x_i, β_1 et β_2. Vérifiez que $\widetilde{\beta_2}$ est un estimateur sans biais de β_2 lorsque la valeur de la constante au sein de la population (β_1) est nulle. Existe-t-il d'autres cas pour lesquels $\widetilde{\beta_2}$ est sans biais ?
   b) Calculez la variance de $\widetilde{\beta_2}$.
   c) Montrez que $V\left(\widetilde{\beta_2}\right) \leq V\left(\widehat{\beta_2}\right)$.
   d) Expliquez le compromis qui existe entre le biais et la variance lorsqu'il s'agit de faire un choix entre $\widehat{\beta_2}$ et $\widetilde{\beta_2}$.

42. Supposons que les variables $y_i, x_{1,i}$, et $x_{2,i}$ satisfont les hypothèses des moindres carrés et que, pour toutes les observations $V\left(\varepsilon_i | x_{1,i}, x_{2,i}\right) = 4$ et $V\left(x_{1,i}\right) = 6$. L'échantillon aléatoire comprend 400 observations.

   a) Supposons que x_1 et x_2 soient non corrélés. Calculez la variance de $\hat{\beta}_1$? Pour cela démontrez que :

$$V\left(\widehat{\beta_1}\right) = \frac{1}{N} \frac{1}{1 - \left(\rho_{x_1,x_2}\right)^2} \frac{{\sigma_\varepsilon}^2}{{\sigma_{x_1}}^2}$$

   où ρ_{x_1,x_2} est la corrélation théorique entre x_1 et x_2, ${\sigma_{x_1}}^2$ est la variance théorique de x_1.

   b) Supposons maintenant que la corrélation entre x_1 et x_2 soit égale à 0.5 : $\rho_{x_1,x_2} = 0.5$. Calculez la variance de $\widehat{\beta_1}$?

   c) Commentez l'affirmation : « *si x_1 et x_2 sont corrélés, la variance de $\widehat{\beta_1}$ est plus grande que dans le cas contraire* » ?

[Stock et Watson (2011), exercice 3.8]

43. Considérons le modèle de régression sans constante :

$$y_i = \beta_1 x_{1,i} + \beta_2 x_{2,i} + \varepsilon_i, \qquad i = 1,2, \dots, N.$$

Répondez aux questions suivantes :

a) Précisez la fonction des moindres carrés, minimisée par les MCO ?
b) Calculez les conditions du premier ordre et donnez les équations normales ?
c) Supposons que $\sum_{i=1}^{N} x_{1,i} x_{2,i} = 0$, montrez qu'on peut alors écrire :

$$\widehat{\beta_1} = \frac{\sum_{i=1}^{N} x_{1,i} y_i}{\sum_{i=1}^{N} {x_{1,i}}^2} \quad et \quad \widehat{\beta_2} = \frac{\sum_{i=1}^{N} x_{2,i} y_i}{\sum_{i=1}^{N} {x_{2,i}}^2}.$$

d) Supposons que $\sum_{i=1}^{N} x_{1,i} x_{2,i} \neq 0$, établissez une expression de $\widehat{\beta_1}$ en fonction des variables $y_i, x_{1,i}$, et $x_{2,i}$?
e) Supposons que le modèle comporte une constante, et s'écrit donc : $y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \varepsilon_i$. Démontrez que les estimateurs des moindres carrés satisfont : $\widehat{\beta_0} = \bar{y} - \widehat{\beta_1}\overline{x_1} - \widehat{\beta_2}\overline{x_2}$?
f) De la même manière que dans (e), supposons que le modèle comporte une constante et que $\sum_{i=1}^{N}\left(x_{1,i} - \overline{x_1}\right)\left(x_{2,i} - \overline{x_2}\right) = 0$. Montrez que

$$\widehat{\beta_1} = \frac{\sum_{i=1}^{N}\left(x_{1,i} - \overline{x_1}\right)(y_i - \bar{y})}{\sum_{i=1}^{N}\left(x_{1,i} - \overline{x_1}\right)^2}$$

g) Que se passe-t-il si la variable x_2 est omise dans la régression ? Comparez les estimateurs de β_1 des MCO correspondants aux régressions avec et sans la variable x_2 ?

[Stock et Watson (2011), exercice 3.9]

44. Dans une régression simple, $\widehat{\beta_1}$ et $\widehat{\beta_2}$ sont les estimateurs de l'ordonnée à l'origine et de la pente de la droite de régression. Soit $\bar{\varepsilon}$ la moyenne des erreurs (et non des résidus !).

a) Montrez que $\widehat{\beta_2}$ peut s'écrire sous la forme $\widehat{\beta_2} = \beta_2 + \sum_{i=1}^{N} w_i \varepsilon_i$ où $w_i = d_i/SCT$ et $d_i = x_i - \bar{x}$.
b) En partant du point (i), montrez que $\sum_i w_i = 0$ et que $\widehat{\beta_2}$ et $\bar{\varepsilon}$ ne sont pas corrélés.
c) Montrez que $\widehat{\beta_1}$ peut s'écrire sous la forme : $\widehat{\beta_1} = \beta_1 + \bar{\varepsilon} - \left(\widehat{\beta_1} - \beta_1\right)\bar{x}$.
d) Utilisez les points (ii) et (iii) pour montrer que $V\left(\widehat{\beta_1}\right) = \frac{\sigma^2}{N} + \frac{\sigma^2 \bar{x}^2}{SCT_x}$.
e) Simplifiez l'expression précédente pour aboutir à l'équation :

$$V\left(\widehat{\beta_1}\right) = \frac{\sigma^2}{N} \frac{\sum_i x_i^2}{\sum_i (x_i - \bar{x})^2}$$

45. Supposons que nous avons deux estimateurs $\left(\widehat{\theta_1} \text{ et } \widehat{\theta_2}\right)$ d'un paramètre θ.Ceux-ci sont sans biais, et indépendants l'un de l'autre. Leur variance est respectivement v_1 et v_2. Quelle est la combinaison linéaire des deux estimateurs : $\hat{\theta} = c_1\widehat{\theta_1} + c_2\widehat{\theta_2}$ qui est l'estimateur sans biais de variance minimale de θ ?
[Greene (2018), exercice 4.1]

46. Supposons un modèle de régression linéaire classique, mais que la vraie valeur de la constante est zéro.
Comparez la variance de l'estimateur MCO des paramètres de pente calculé sans la constante avec celle de l'estimateur MCO des paramètres de pente calculé avec cette constante inutile.
[Greene (2018), exercice 4.3]

47. Considérons la régression multiple de $\boldsymbol{y}$ sur K variables explicatives $\boldsymbol{X}$ et une variable explicative additionnelle $\boldsymbol{z}$.
Sous les hypothèses H1 à H4 du modèle de régression linéaire classique, prouvez que la vraie variance de l'estimateur des MCO des paramètres de X est plus grande quand la variable z est incluse dans la régression que dans le cas où elle n'est incluse dans la régression. (on suppose que le coefficient de z n'est pas zéro)
[Greene (2018), exercice 4.8]

48. Un chercheur dispose de deux échantillons indépendants des variables (Y_i, X_i) qui désignent respectivement les salaires et le nombre d'années d'étude pour les hommes et pour les femmes.
Pour les hommes, (1er échantillon), considérons la régression :

$$Y_{M,i} = \beta_{1,M} + \beta_{2,M}X_{M,i} + \varepsilon_{M,i}$$

Pour les femmes, (2ème échantillon), considérons la régression :

$$Y_{F,i} = \beta_{1,F} + \beta_{2,F}X_{M,i} + \varepsilon_{F,i}$$

$s_{\widehat{\beta_{2,M}}}$ et $s_{\widehat{\beta_{2,F}}}$ représentent respectivement l'écart-type estimé du paramètre de la variable X dans les deux régressions. Montrez que l'écart-type estimé de $\widehat{\beta_{2,M}} - \widehat{\beta_{2,F}}$ s'écrit :

$$s_{(\widehat{\beta_{2,M}}-\widehat{\beta_{2,F}})} = \sqrt{\left(s_{\widehat{\beta_{2,M}}}\right)^2 + \left(s_{\widehat{\beta_{2,F}}}\right)^2}$$

[Stock et Watson (2011), exercice 2.15]

49. Si les contraintes sont correctes, démontrez l'estimateur des moindres carrés contraint est sans biais. Quelles sont les hypothèses que vous devez supposer ? Montrez que l'estimateur des MCO est aussi sans biais.
Dérivez la matrice de variance – covariance de l'estimateur des MCC. Montrez qu'elle est plus petite (au sens matriciel) que celle des MCO. Interprétez ce résultat.

50. Donnez une expression du biais de l'estimateur des MCC si la contrainte n'est pas correcte ?

51. On sait que la somme des carrés des résidus des MCC est supérieure à celle des MCO. Donnez la condition pour que l'estimateur de la variance de l'erreur pour les MCC (σ^{*2}) soit supérieure à l'estimateur de la variance de l'erreur des MCO ($\hat{\sigma}^2$).

52. Considérons le modèle de régression multiple avec 3 variables explicatives sous les hypothèses usuelles du modèle linéaire classique :

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

Vous souhaitez tester l'hypothèse nulle $H_0: \beta_1 - 3\beta_2 = 1$.

a) Soit $\widehat{\beta_1}$ et $\widehat{\beta_2}$ les estimateurs des MCO des paramètres β_1 et β_2 respectivement. Ecrivez $Var\left(\widehat{\beta_1} - 3\widehat{\beta_2}\right)$ en fonction des variances et covariances des paramètres estimés $\widehat{\beta_1}$ et $\widehat{\beta_2}$. Quelle est l'expression de l'écart-type de $\widehat{\beta_1} - 3\widehat{\beta_2}$?
b) Ecrivez la statistique t permettant de tester $H_0: \beta_1 - 3\beta_2 = 1$.
c) On définit $\theta_1 = \beta_1 - 3\beta_2$. Réécrivez l'équation de départ de sorte à faire apparaître les paramètres $\beta_0, \theta_1, \beta_2$ et β_3 dans l'équation et ainsi obtenir directement l'estimateur de $\widehat{\theta_1}$.

d) On définit aussi $\theta_2 = \beta_3 + 2\beta_2$. Quels sont les variances de $\widehat{\theta_1}$ et $\widehat{\theta_2}$, et la covariance entre $\widehat{\theta_1}$ et $\widehat{\theta_2}$ en fonction des variances et covariances des $\widehat{\beta_J}$.

53. Pour un nombre d'observation (N) et un nombre de paramètres donnés (K), quel est le niveau du R^2 qui assure le rejet de l'hypothèse nulle que tous les paramètres de pente soient zéros ? Utilisez pour cela le test F de significativité globale d'une régression. Quel est ce niveau de R^2 pour $K = 5$ et $N = 20, 50, 100$ ou 1000 ?

Chapitre 3 : Les propriétés asymptotiques de l'estimateur des MCO

54. Supposons que $\sqrt{N}\left(\widehat{\theta_N} - \theta\right) \overset{d}{\rightarrow} N(0\,, \sigma^2)$. Prouvez que $\widehat{\theta_N} \overset{p}{\rightarrow} \theta$.
    [Hayashi, 2000, Section II.1, Exercice 4]

55. Soit $\{z_i\}$ une séquence de variables aléatoires indépendantes et identiquement distribuées avec $E(z_i) = \mu \neq 0$ et $V(z_i) = \sigma^2$. La moyenne d'échantillon est $\overline{z_N} = \frac{1}{N}\sum_{i=1}^{N} z_i$. Démontrez que

    $$\sqrt{N}\left(\frac{1}{\overline{z_N}} - \frac{1}{\mu}\right) \overset{d}{\rightarrow} N\left(0, \frac{\sigma^2}{\mu^4}\right)$$

    [Hayashi, 2000, Section II.1, Exercice 5]

56. Soit un vecteur $K \times 1$ de variables aléatoires $\boldsymbol{z}_t$ stationnaires en covariance avec une matrice des autocovariances $\boldsymbol{\Gamma}_s = Cov(\boldsymbol{z}_t, \boldsymbol{z}_{t-s})$. Montrez que $\boldsymbol{\Gamma}_s = \boldsymbol{\Gamma}'_{-s}$
    [Hayashi, 2000, Section II.2, Exercice 1]

57. Supposons que le processus $\{x_t\}$ soit une martingale par rapport à $\{\boldsymbol{z}_t\}$. Montrez que $E(x_{t+s}|\boldsymbol{z}_{t-1}, \boldsymbol{z}_{t-1}, \ldots \boldsymbol{z}_1) = x_{t-1}$ et que $E(x_{t+s+1} - x_{t+s}|\boldsymbol{z}_{t-1}, \boldsymbol{z}_{t-1}, \ldots \boldsymbol{z}_1) = 0$ pour $s = 0,1, \ldots$
    [Hayashi, 2000, Section II.2, Exercice 3]

58. Soit $\{x_i\}$ une séquence de nombres réels et $\{\varepsilon_i\}$ une séquence de variables aléatoires indépendantes et identiquement distribuées d'espérance nulle et de variance finie. Est-ce que la séquence $\{x_i\varepsilon_i\}$ est indépendante et identiquement distribuée ? A-t-elle des autocorrélations nulles ? Est-elle une martingale en différence ? Est-elle stationnaire ?
    [Hayashi, 2000, Section II.2, Exercice 4]

59. Définissez une marche aléatoire. Quelle est son espérance, sa variance et ses autocorrélations. Montrez que cette marche aléatoire n'est pas stationnaire.
[Hayashi, 2000, Section II.2, Exercice 5]

60. Supposons le modèle de régression sans variables explicatives : $y_i = \mu + \varepsilon_i$ avec $\varepsilon_i \approx i.i.d. N(0, \sigma^2)$. Prouvez que la moyenne d'échantillon $\hat{\mu} = \bar{y} = \frac{1}{N}\sum_i y_i$ est un estimateur convergent et asymptotiquement normal. Maintenant considérez l'estimateur alternatif : $\tilde{\mu} = \sum_i w_i y_i$ avec $w_i = i/\sum_i i$. Montrez que cet estimateur est convergent et calculez se variance asymptotique. Pour rappel : $\sum_i i = n(n+1)/2$ et $\sum_i i^2 = n(n+1)(2n+1)/6$.
[Greene (2018), exercice 4.8]

61. Soit $s_{\widehat{\beta_k}}$ l'écart-type estimé d'un paramètre estimé par MCO. Montrez que $s_{\widehat{\beta_k}} \xrightarrow{p} 0$
[Hayashi, 2000, Section II.4, Exercice 1]

62. Supposons par simplicité que $K = 1$ et que $\hat{\beta}$ est l'estimateur MCO de β. L'écart-type de $\hat{\beta}$ est $\sqrt{Avar(\hat{\beta})/N}$. On définit $\lambda = -\log \beta$ qui est estimé par $\hat{\lambda} = -\log \hat{\beta}$. Vérifiez que l'écart-type de$\hat{\lambda}$ est alors $(1/\hat{\beta})\sqrt{Avar(\hat{\beta})/N}$.
[Hayashi, 2000, Section II.4, Exercice 2]

63. Il n'y a pas une manière unique d'écrire les J hypothèses linéaires $\boldsymbol{R\beta} = \boldsymbol{r}$. On peut représenter les mêmes restrictions par $\widetilde{\boldsymbol{R}}\boldsymbol{\beta} = \tilde{\boldsymbol{r}}$ en prémultipliant $\boldsymbol{R}$ et $\boldsymbol{r}$ par une matrice $\boldsymbol{F}$ de dimension $J \times J$ non singulière : $\widetilde{\boldsymbol{R}} = \boldsymbol{FR}$ et $\tilde{\boldsymbol{r}} = \boldsymbol{Fr}$. Est-ce que les choix de $\boldsymbol{R}$ ou de $\widetilde{\boldsymbol{R}}$ modifient la valeur numérique du test de Wald, sa distribution asymptotique, sa distribution en échantillon fini ?
[Hayashi, 2000, Section II.4, Exercice 3]

64. On veut tester la relation non linéaire entre les paramètres d'un modèle : $\beta_1\beta_2 = 1$ avec le test de Wald. Montrez que la statistique de Wald de l'hypothèse nulle $H_0: \beta_1\beta_2 - 1 = 0$ est mathématiquement différente de la statistique de test de l'hypothèse nulle $H_0: \beta_1 - (1/\beta_2) = 0$. Est-ce que ces deux statistiques sont asymptotiquement équivalentes ?

65. On veut tester la relation non linéaire entre les paramètres d'un modèle : $\beta_1\beta_2 = \beta_3$ avec le test de Wald. Montrez que la statistique de Wald de l'hypothèse nulle $H_0: \beta_1\beta_2 - \beta_3 = 0$ est mathématiquement différente de la statistique de test de l'hypothèse nulle $H_0: \beta_1 - (\beta_3/\beta_2) = 0$. Est-ce que ces deux statistiques sont asymptotiquement équivalentes ?

66. Montrez que la vraie variance de l'estimateur des MCO est plus grande que la vraie variance de l'estimateur des MCG.

67. Parmi les propositions suivantes, quelles sont celles susceptibles d'être causées par la présence d'hétéroscédasticité ? Expliquez.
    a) Les estimateurs des MCO, $\widehat{\beta_k}$, ne sont pas convergents.
    b) La statistique *F* habituelle ne suit plus une distribution *F* de Fisher.
    c) Les estimateurs des MCO ne sont plus les meilleurs estimateurs linéaires sans biais.

68. Avec les hypothèses 2.1 à 2.5, la variance asymptotique de l'estimateur des MCO est $Avar(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{S}\boldsymbol{\Sigma}_{xx}^{-1}$ avec $\boldsymbol{S} = E(\varepsilon_i^2\boldsymbol{x}_i\boldsymbol{x}_i')$. Est-elle estimée de manière convergente par $\widehat{Avar(\widehat{\boldsymbol{\beta}})} = N\widehat{\sigma^2}\mathbf{S}_{xx}^{-1}$ avec $\boldsymbol{S}_{xx} = \left(\frac{1}{N}\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$ sans l'hypothèse (2.7) d'homoscédasticité conditionnelle ? Est-ce que la statistique de test $t_{\widehat{\beta_k}} = \widehat{\beta_k}/s_{\widehat{\beta_k}}$ est asymptotiquement distribué selon une loi normale standard.
[Hayashi, 2000, Section II.6, Exercice 1]

69. Sans l'hypothèse (2.7) d'homoscédasticité conditionnelle, est-ce que la statistique de test de J contraintes $(SCR_C - SCR)/\widehat{\sigma^2}$ est asymptotiquement distribué selon une loi du Khi-deux avec J degré de liberté.
[Hayashi, 2000, Section II.6, Exercice 4]

70. Dans une régression avec une constante, considérez l'hypothèse nulle que tous les $K-1$ paramètres de pente soient nuls. Montrez que $NR^2 \xrightarrow{d} \chi^2(K-1)$ sous les hypothèses 2.1 à 2.5 et l'hypothèse (2.7) d'homoscédasticité conditionnelle. Pourquoi ne peut-on pas utiliser cette expression lorsqu'il y a de l'hétéroscédasticité conditionnelle.
[Hayashi, 2000, Section II.6, Exercice 5]

71. Quelle est la matrice de covariance, $Cov\left(\widehat{\boldsymbol{\beta}_{MCG}}, \widehat{\boldsymbol{\beta}_{MCG}} - \widehat{\boldsymbol{\beta}_{MCO}}\right)$, entre l'estimateur des *MCG* : $\widehat{\boldsymbol{\beta}_{MCG}} = (\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{y}$ et sa différence avec l'estimateur des *MCO* : $\widehat{\boldsymbol{\beta}_{MCO}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$ **.**
[Greene (2018), exercice 9.1]

72. *[Théorème de Kruskal]* Dans un modèle de régression par *MCG*, si on peut écrire que les K colonnes de $\boldsymbol{X}$ sont les vecteurs caractéristiques (*eigenvector*) de $\boldsymbol{\Omega} = E(\boldsymbol{\varepsilon\varepsilon}')$, alors les estimateurs des *MCG* et des *MCO* sont identiques.
[Greene (2018), exercice 9.5]

73. Dans un modèle de régression par *MCG* avec la matrice $\boldsymbol{\Omega} = E(\boldsymbol{\varepsilon\varepsilon}')$ connue,
    a) Quelle est la matrice de covariance entre les estimateurs *MCG* ($\widehat{\boldsymbol{\beta}_{MCG}}$) et *MCO* ($\widehat{\boldsymbol{\beta}_{MCO}}$) ?
    b) Quelle est la matrice de variance-covariance des résidus des *MCO* ($\boldsymbol{e}_{MCO} = \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}_{MCO}}$) ?
    c) Quelle est la matrice de variance-covariance des résidus des *MCG* ($\boldsymbol{e}_{MCG} = \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}_{MCG}}$)
    d) Quelle est la matrice de covariance entre les résidus des *MCG* ($\boldsymbol{e}_{MCG}$) et les résidus des *MCO* ($\boldsymbol{e}_{MCO}$) ?

[Greene (2018), exercice 9.6]

74. Considérons le modèle linéaire suivant pour expliquer la consommation mensuelle de bière :

$$bière = \beta_1 + \beta_2 revenu + \beta_3 prix + \beta_4 etude + \beta_5 femme + \varepsilon$$

$$\text{avec : } \begin{cases} E(u|revenu, prix, etude, femme) = 0 \\ V(u|revenu, prix, etude, femme) = \sigma^2 revenu^2 \end{cases}$$

Ecrivez le modèle transformé avec des erreurs homoscédastiques. Interprétez ce modèle.

[Wooldridge (2009), exercice 8.2]

75. Considérons un modèle appliqué à un ensemble d'employés :

$$y_{i,e} = \beta_0 + \beta_1 x_{1,i,e} + \beta_2 x_{2,i,e} e + \cdots + \beta_4 x_{k,i,e} + v_{i,e} + f_i$$

où la variable non observée f_i capture un « effet d'entreprise », c'est-à-dire l'effet que les caractéristiques propres de l'entreprise *i* peuvent avoir sur la variable dépendante. Le terme d'erreur $v_{i,e}$ est spécifique à l'employé *e* de l'entreprise *i*. L'erreur composite $u_{i,e} = v_{i,e} + f_i$ est celle de l'équation (8.28).

a) Supposons que $Var(f_i) = \sigma_f^2$,$Var(v_{i,e}) = \sigma_v^2$ et que $v_{i,e}$ et f_i ne sont pas corrélés. Montrez alors que $Var(u_{i,e}) = \sigma_f^2 + \sigma_v^2$. Appelez cette variance σ^2.
b) Supposons maintenant que $v_{i,e}$ et $v_{i,g}$ ne sont pas corrélés pour $e \neq g$. Montrez que $v(u_{i,e}\ ,\ u_{i,g}) = \sigma_f^2$.
c) Soit la moyenne des erreurs composites au sein d'une entreprise : $\bar{u_\imath} = m_i^{-1} \sum_{e=1}^{m_i} u_{i,e}$ avec m_i le nombre d'employés d'une entreprise *i*. Montrez que $Var(\bar{u_\imath}) = \sigma_f^2 + \sigma_v^2/m_i$.
d) Discutez l'intérêt du point (iii) pour l'application de la méthode des MCP à des données agrégées au niveau de l'entreprise, où le poids utilisé pour l'observation *i* est fonction de la taille de l'entreprise.

[voir Wooldridge (2009), exercice 8.7]

76. Démontrez que la statistique de Lung-Box est supérieure à la statistique de Box-Pierce. Montrez que ces deux statistiques de test sont asymptotiquement équivalentes. Qu'en concluez-vous ?

77. Supposons un processus stochastique $\{y_t\}$ qui est généré par le modèle : $y_t = z + \varepsilon_t$ pour $t = 1,2, \ldots$ avec $\{\varepsilon_t\}$ une séquence *i.i.d.* d'espérance nulle et de variance σ_ε^2. La variable aléatoire z est invariante dans le temps avec une espérance nulle et une variance σ_z^2. On suppose que chaque ε_t n'est pas corrélé avec z.

    a) Trouvez l'espérance et la variance de y_t. Est-ce qu'elles dépendent du temps ?
    b) Donnez la covariance : $Cov(y_t, y_{t+h})$ pour tout t et h. Est-ce que le processus $\{y_t\}$ est stationnaire en covariance ?
    c) Montrez que les autocorrélations sont : $Corr(y_t, y_{t+h}) = \sigma_z^2/(\sigma_z^2 + \sigma_\varepsilon^2)$ pour tout t et h.
    d) Est-ce que y_t est asymptotiquement non corrélé ? Expliquez.

    [voir Wooldridge (2009), exercice 11.3]

78. Considérez le modèle : $y_t = \boldsymbol{x}_t'\boldsymbol{\beta} + u_t$ avec $u_t = \rho u_{t-1} + \varepsilon_t$ où ε_t est un bruit blanc gaussien. Comparez les autocorrélations de u_t dans le modèle original, avec celles de v_t dans le modèle en différence première : $(y_t - y_{t-1}) = (\boldsymbol{x}_t - \boldsymbol{x}_{t-1})'\boldsymbol{\beta} + (u_t - u_{t-1})$ ou $\Delta y_t = \Delta \boldsymbol{x}_t{}'\boldsymbol{\beta} + v_t$ avec $v_t = u_t - u_{t-1}$.
    Est-ce que la transformation en différence première réduit l'autocorrélation ?
    [voir Greene (2018), exercice 20.1]

Chapitre 4 : L'estimateur de la Méthode des Moments Généralisée (GMM)

79. Soit un modèle d'offre – demande classique de la forme :

$$\begin{cases} q_i = \alpha_0 + \alpha_1 p_i + u_i & \rightarrow \text{ équation de demande} \\ q_i = \beta_0 + \beta_1 p_i + v_i & \rightarrow \text{ équation d'offre} \end{cases}$$

a) Supposons que la covariance entre les deux erreurs n'est pas nulle : $Cov(u_i, v_i) \neq 0$. est-ce que le prix et le choc de demande (u_i) sont positivement corrélés quand $\alpha_1 < 0$ et $\beta_1 > 0$?

b) On a vu que l'estimateur des MCO du paramètres du prix dans une régression de la quantité sur une constante et la prix est biaisé pour α_1. Est-ce que le paramètre estimé de la constante est aussi biaisé pour α_0 ?

c) Démontrez que l'estimateur du paramètre du prix dans une régression de la quantité sur une constante et le prix ($q_i = \gamma_0 + \gamma_1 p_i + \eta_i$) a pour limite en probabilité :

$$\text{plim } \widehat{\gamma_1} = \frac{\alpha_1 V(v_i) + \beta_1 V(u_i)}{V(v_i) + V(u_i)}$$

[Hayashi (2000), Exercices 3.1.(1), (2) et (3)]

80. Sous quelles hypothèses l'estimateur par variables instrumentales est convergent ? Démontrez la convergence de cet estimateur. Donnez la distribution asymptotique de l'estimateur VI.

81. Montrez que la régression par MCO est un cas particulier de l'estimation par variables instrumentales.

82. Soit $\boldsymbol{\Sigma}_{\boldsymbol{xz}} = E(\boldsymbol{x}_i \boldsymbol{z}_i')$ de dimension $K \times L$ et $\boldsymbol{\Sigma}_{\boldsymbol{xy}} = E(\boldsymbol{x}_i y_i)$ de dimension $K \times 1$. Montrez que :

$$\text{rang}(\boldsymbol{\Sigma}_{\boldsymbol{xz}}) = \text{rang}\big(\boldsymbol{\Sigma}_{\boldsymbol{xz}} \ \vdots \ \boldsymbol{\Sigma}_{\boldsymbol{xy}}\big)$$

[Hayashi (2000), Exercices 3.3(5)]

83. Si on ajoute une autre variable instrumentale (ξ_i) aux instruments ($\boldsymbol{x}_i$). Bien que cette variable supplémentaire soit prédéterminée, elle n'est pas reliée aux régresseurs : $E(\xi_i z_{i,l}) = 0$ pour tout $l = 1,2,\dots,L$. Est-ce que la condition de rang est toujours satisfaite ?
[Hayashi (2000), Exercices 3.3(6)]

84. Soit la matrice $\boldsymbol{A}$ de dimension $q \times K$ de rang plein ligne avec $q \leq K$ telle que $\boldsymbol{A\Sigma_{xz}}$ est de rang plein colonne. Notons $\widehat{\boldsymbol{x}_\iota} = \boldsymbol{A}\boldsymbol{x}_i$. Vérifiez que les hypothèses 3.3, 3.4 et 3.5 tiennent pour les instruments transformés $\widehat{\boldsymbol{x}_\iota}$ si elles sont satisfaites pour les variables instrumentales $\boldsymbol{x}_i$.
[Hayashi (2000), Exercices 3.3(8)]

85. Si l'équation est juste identifiée, quelle est la valeur minimisée de $J(\widetilde{\boldsymbol{\delta}}, \widehat{\boldsymbol{W}})$?
[Hayashi (2000), Exercices 3.4(2)]

86. Soit un modèle de régression simple $y_t = \alpha + \beta z_t + \varepsilon_t$ avec $\varepsilon_t \approx i.i.d.(0, \sigma_\varepsilon^2)$. On suppose que $E(z_t \varepsilon_t) \neq 0$, mais qu'on a une variable instrumentale x_t corrélée avec z_t telle que $E(x_t \varepsilon_t) = 0$.
Montrez que l'estimateur par variable instrumentale de β est équivalent à l'estimateur par *MCO* de β dans modèle : $y_t = \alpha + \beta z_t + \gamma v_t + \varepsilon_t$ où v_t est le résidus de la régression de première étape : $z_t = \delta + \theta x_t + \nu_t$.
[Proposition de Durbin (1954) et Wu (1973)]

87. Supposons que l'on recueille par une enquête des données sur la consommation de vin des ménages, ainsi que leurs revenus. On veut estimer l'élasticité de la consommation de vin au revenu. Mais ces deux variables sont mal renseignées, ou avec des erreurs.

    a) Quelle est la conséquence des erreurs de mesure sur la variable dépendante (consommation de vin) et sur la variable explicative (le revenu) ?
    b) Un économètre propose d'utiliser la variable instrumentale : montant total des chèques émis par le ménage. Pensez-vous que cette variable est un bon instrument ? Pourquoi ?

88. Dans un modèle de détermination des salaires, on explique souvent le log ($salaire$) par l'éducation mesurée par le nombre d'années d'étude ($etude$) et par d'autres variables explicatives $\boldsymbol{X}$. Soit le modèle :

$$\log(salaire) = \beta_1 + \beta_2 etude + \boldsymbol{X\gamma} + \varepsilon$$

Or le nombre d'année d'étude ne mesure qu'imparfaitement la qualité de la formation suivie par la personne. Quel est alors la conséquence de prendre la variable $etude$ dans cette régression au lieu de la qualité de la formation ?

89. Dans un modèle où la variable explicative est endogène, l'estimateur des MCO est biaisé et non convergent : plim $\widehat{\boldsymbol{\beta}_{MCO}} = \boldsymbol{\beta} + \boldsymbol{Q}^{-1}\boldsymbol{\gamma} = \boldsymbol{\theta}$ avec $\boldsymbol{Q} = \text{plim}(\boldsymbol{X}'\boldsymbol{X}/N)$ et $\boldsymbol{\gamma} = \text{plim}(\boldsymbol{X}'\boldsymbol{\varepsilon}/N)$.

   a) Dérivez la matrice de variance-covariance asymptotique de l'estimateur $\widehat{\boldsymbol{\beta}_{MCO}}$.

   b) Montrez que l'estimateur $\widehat{\boldsymbol{\beta}_{MCO}}$ est asymptotiquement normalement distribué.

[Greene (2018) : Exercice 8.1]

90. Supposons deux choix de matrice de pondération $\widehat{\boldsymbol{W}_{\mathbf{1}}}$ et $\widehat{\boldsymbol{W}_{\mathbf{2}}}$ tel que $\widehat{\boldsymbol{W}_{\mathbf{1}}} - \widehat{\boldsymbol{W}_{\mathbf{2}}} \overset{p}{\to} \mathbf{0}$. Démontrez que :

$$\sqrt{N}\,\widehat{\boldsymbol{\delta}}\big(\widehat{\boldsymbol{W}_{\mathbf{1}}}\big) - \sqrt{N}\,\widehat{\boldsymbol{\delta}}\big(\widehat{\boldsymbol{W}_{\mathbf{2}}}\big) \overset{p}{\to} \mathbf{0}$$

[Hayashi (2000), Exercices 3.5(3)]

91. Démontrez que :

$$J\big(\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}}^{-1}), \widehat{\boldsymbol{S}}^{-1}\big) = N\boldsymbol{s}_{\boldsymbol{xy}}'\widehat{\boldsymbol{S}}^{-1}\left(\boldsymbol{s}_{\boldsymbol{xy}} - \boldsymbol{S}_{\boldsymbol{xz}}\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}}^{-1})\right)$$

[Hayashi (2000), Exercices 3.6(2)]

92. En cas d'homoscédasticité conditionnelle, la statistique de test de Sargan est :

$$Sargan = N\frac{\big(\boldsymbol{s}_{\boldsymbol{xy}} - \boldsymbol{S}_{\boldsymbol{xz}}\widehat{\boldsymbol{\delta}_{DMC}}\big)'\boldsymbol{S}_{\boldsymbol{xx}}^{-1}\big(\boldsymbol{s}_{\boldsymbol{xy}} - \boldsymbol{S}_{\boldsymbol{xz}}\widehat{\boldsymbol{\delta}_{DMC}}\big)}{\widehat{\sigma^2}}$$

Démontrez que cette statistique est égale à $N \times R_{nc}^2$, le coefficient de détermination non centré d'une régression de $\boldsymbol{e}_{DMC} = \boldsymbol{y} - \boldsymbol{Z}\widehat{\boldsymbol{\delta}_{DMC}}$ sur $\boldsymbol{X}$.
[Hayashi (2000), Exercices 3.8(6)]

93. Considérez le modèle de régression : $y_i = \alpha + \beta x_i^* + \varepsilon_i$. Mais la variable x_i^* n'est pas observée, on observe seulement une variable avec erreurs : $x_i = x_i^* + \nu_i$. Cependant on dispose d'une indicatrice permettant de classer les observations : $d_i = 1$ ou 0 .

a) Montrez que l'estimateur par variables instrumentales utilisant cette indicatrice comme variable instrumentale est :

$$\widehat{\beta_{VI}} = \frac{\overline{y_{d=1}} - \overline{y_{d=0}}}{\overline{x_{d=1}} - \overline{x_{d=0}}}$$

où $\overline{y_{d=1}}$ est la moyenne de y pour les observations avec $d_i = 1$, et ainsi de suite. Interprétez cette expression.

b) Quelle est la condition pour laquelle cet estimateur est calculable, interprétez cette condition ?

c) Quel est l'expression pour l'estimateur $\widehat{\alpha_{VI}}$?

[voir Greene (2018) : Exemple 8.2 et Exercice 8.5, ou Wooldridge(2009) : Exercice 15.3]

94. Considérez le modèle de régression simple sur séries temporelles où la variable explicative est sujette à des erreurs de mesure :

$$y_t = \alpha + \beta x_t^* + \varepsilon_t$$
$$x_t = x_t^* + \nu_t$$

a) Introduisez $x_t^* = x_t - \nu_t$ dans le modèle et montrez que le terme d'erreur de la nouvelle équation (u_t) est négativement corrélé avec x_t si $\beta > 0$. Quelle est l'implication pour l'estimateur des MCO de β dans la régression de y sur x ?

b) En plus des précédentes hypothèses, on suppose que ε_t et ν_t ne sont pas corrélés avec toutes les valeurs passées de x_t^* et de ν_t, et en particulier avec x_{t-1}^* et ν_{t-1}. Montrez alors que $E(x_{t-1}u_t) = 0$.

c) Est-ce que x_t et x_{t-1} sont probablement corrélés ? Expliquez.

d) Donnez une stratégie pour estimer de manière convergente α et β dans ce modèle.

[Wooldridge(2009) : Exercice 15.11]

Chapitre 5 : La méthode du maximum de vraisemblance (MV)

95. Considérons un échantillon aléatoire de 10 observations :

5 ; 0 ; 1 ; 1 ; 0 ; 3 ; 2 ; 3 ; 4 ; 1

provenant d'une distribution de Poisson. La fonction de densité de chaque observation est :

$$f(y_i|\theta) = \frac{e^{-\theta}\theta^{y_i}}{y_i!} \quad \text{avec } y_i = 0,1,2,\ldots \text{ et } \theta > 0$$

Pour une distribution de Poisson : $E(y_i) = V(y_i) = \theta > 0$.

a) Quelle est la vraisemblance de cet échantillon en fonction du paramètre θ.

b) Donnez en général pour des observations suivant une loi de Poisson, la log-vraisemblance et l'équation de vraisemblance pour l'estimateur du maximum de vraisemblance.

c) Calculez l'estimateur du maximum de vraisemblance de θ. Vérifiez que vous avez bien un maximum de la log-vraisemblance.

d) Donnez la probabilité estimée des valeurs de 0 à 5.

[voir Greene (2018), Chapitre XIV.3]

96. Dans un modèle Probit, quel est l'effet marginal d'une variable explicative sur la probabilité de choix ?

a) Si la variable explicative est continue

b) Si la variable explicative est une indicatrice.

97. On a un échantillon de disques durs d'ordinateurs dont on a mesuré la durée de vie y_i avant de tomber en panne. On va supposer que cette variable dépendante y_i est distribuée selon une loi exponentielle. Celle-ci permet de mesurer la durée de vie d'un phénomène dont le taux de panne à un moment est constant dans le temps. Sa fonction de densité est :

$$f(y_i) = \frac{1}{\mu_i} \exp\left(\frac{-y_i}{\mu_i}\right) \quad \text{pour } y_i > 0$$

où $\mu_i = E(y_i) > 0$ est l'espérance de y_i. Par définition, sa variance est $V(y_i) = \mu_i^2$. On suppose que $\mu_i = \exp(x_i'\beta) > 0$ pour assurer la positivité.

a) Donnez la fonction de log-vraisemblance pour un échantillon aléatoire de N observations indépendantes.

b) Donnez les équations de vraisemblance.

c) Peut-on donner une expression analytique pour l'estimateur du maximum de vraisemblance de β ?

d) Donnez une expression pour la matrice de variance-covariance asymptotique de l'estimateur du maximum de vraisemblance.

98. Supposons que la variable x a une distribution de Weibull telle que la fonction de densité est : $f(x) = \alpha\beta x^{\beta-1} \exp\left(-\alpha x^{\beta}\right)$ pour $x \geq 0$ et $\alpha, \beta > 0$.

a) Donnez la fonction de log-vraisemblance pour un échantillon aléatoire de N observations.

b) Donnez les équations de vraisemblance pour l'estimation par maximum de vraisemblance de α et β.

c) Obtenez la matrice des dérivées secondes de la log-vraisemblance par rapport à α et β. Les espérances exactes des éléments impliquant β impliquent des dérivées de la fonction Gamma. Ce qui est assez difficile à trouver analytiquement. Votre résultat exact donne un estimateur empirique. Comment estimeriez-vous plus facilement la matrice de covariance asymptotique de l'estimateur MV ?

d) Prouvez que $\alpha\beta Cov\left(\ln x\ , x^{\beta}\right) = 1$.
[Utilisez le fait que l'espérance des dérivées premières de la log-vraisemblance est nulle]

[voir Greene (2018), exercice 14.4]

99. On veut tester les J hypothèses linéaires $\boldsymbol{R\beta} = \boldsymbol{r}$. On peut utiliser un test de Wald de l'estimateur des multiplicateurs de Lagrange des MCC, soit les hypothèses : H_0 : $\boldsymbol{\lambda} = \boldsymbol{0}$. Le test de Wald est : $W = \boldsymbol{\lambda}^{*\prime}\hat{V}(\boldsymbol{\lambda}^*)^{-1}\boldsymbol{\lambda}^*$.

a) Prouvez que ce test dépend du ratio des estimateurs de la variance de l'erreur :

$$W = (N-K)\left[\frac{\boldsymbol{e}^{*\prime}\boldsymbol{e}^*}{\boldsymbol{e}'\boldsymbol{e}} - 1\right]$$

b) Pourquoi le facteur entre crochets est positif ?

c) Prouvez que cette statistique de test est égale à $J \times F$, où F est la statistique de test des restrictions $\boldsymbol{R\beta} = \boldsymbol{r}$ dans le modèle non contraint.

d) Montrez que la statistique de test du multiplicateur de Lagrange peut se réécrire comme

$$LM = \frac{N \times J}{((N-K)/F) + J}$$

[voir Greene (2018), exercice 5.7]

100. Dans un modèle *Logit* ou *Probit* où il n'y a que la constante : $y_i = \beta_1 + \varepsilon_i$. Notez P : le nombre d'observations avec une valeur $y = 1$ parmi les N observations.

a) Donnez l'estimateur de *MV* de la constante dans le modèle *Logit* et *Probit*.

b) Quelle est la valeur maximale de la log-vraisemblance dans ces deux modèles ?

c) Proposez un test de significativité conjointe des paramètres de pente s'il y a d'autres régresseurs dans le modèle : $y_i = \beta_1 + \beta_2 x_2 + \cdots + \beta_K x_K + \varepsilon_i$.

ÉCONOMETRIE THEORIQUE

(M1 MBFA)

Corrigés des Exercices Théoriques

Chapitre 2 : Propriétés en petits échantillons de l'estimateur des MCO

33. Soit deux variables aléatoires x et y avec :

$$E\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix} \quad et \quad V\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix} = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix}$$

a) Montrez que ρ est la corrélation entre x et y ?

Par définition, la corrélation entre les deux variables est :

$$\rho = Corr(x, y) = \frac{Cov(x, y)}{\sqrt{V(x) \times V(y)}}$$

Ici $V(x) = \sigma_x^2, V(y) = \sigma_y^2$, et $Cov(x, y) = \sigma_{xy}$, on aura alors :

$$\rho = \frac{\sigma_{xy}}{\sqrt{\sigma_x^2\sigma_y^2}} = \frac{\sigma_{xy}}{\sigma_x\sigma_y}$$

ou encore : $\sigma_{xy} = \rho\sigma_x\sigma_y$.

b) Quelle est l'espérance et la variance de $3x$?

Espérance : $E(3x) = 3E(x) = 3\mu_x$

Variance : $V(3x) = E[(3x) - E(3x)]^2 = E\big[3\big(x - E(x)\big)\big]^2 = 3^2E\big(x - E(x)\big)^2 = 9V(x) = 9\sigma_x^2$

c) Quelle est l'espérance et la variance de $x - y$?

Espérance : $E(x-y) = E(x) - E(y) = \mu_x - \mu_y$. L'espérance est un opérateur linéaire : l'espérance d'une somme est la somme des espérances.

Variance : en prenant la définition de la variance :

$$V(x-y) = E[(x-y) - E(x-y)]^2 = E\big[\big(x - E(x)\big) - \big(y - E(y)\big)\big]^2$$

qu'on développe :

$$= E\big(x - E(x)\big)^2 + E\big(y - E(y)\big)^2 - 2E\big[\big(x - E(x)\big)\big(y - E(y)\big)\big]$$

$$= V(x) + V(y) - 2Cov(x,y) = \sigma_x^2 + \sigma_y^2 - 2\sigma_{xy}$$

(Attention à ne pas oublier la covariance avec son signe. Même si on a une différence de variables aléatoires, les variances s'additionnent !)

d) Quelle est la covariance et la corrélation entre $(x+y)$ et $(x-y)$?

On a vu plus haut que $V(x-y) = V(x) + V(y) - 2Cov(x,y)$. On peut prouver de la même manière que $V(x+y) = V(x) + V(y) + 2Cov(x,y)$.

Calculons la covariance entre $(x+y)$ et $(x-y)$:

$$Cov(x+y, x-y) = E\big[\big((x+y) - E(x+y)\big)\big((x-y) - E(x-y)\big)\big]$$

$$= E\left[\Big(\big(x - E(x)\big) + \big(y - E(y)\big)\Big)\Big(\big(x - E(x)\big) - \big(y - E(y)\big)\Big)\right]$$

$$= E\big[\big(x - E(x)\big)\big]^2 - E\big[\big(x - E(x)\big)\big(y - E(y)\big)\big] + E\big[\big(y - E(y)\big)\big(x - E(x)\big)\big] - E\big[\big(y - E(y)\big)\big]^2$$

$$= V(x) - V(y)$$

Donc la corrélation entre $(x+y)$ et $(x-y)$ sera :

$$Corr(x+y, x-y) = \frac{Cov(x+y, x-y)}{\sqrt{V(x+y)V(x-y)}}$$

$$= \frac{V(x) - V(y)}{\sqrt{\big(V(x) + V(y) + 2Cov(x,y)\big)\big(V(x) + V(y) - 2Cov(x,y)\big)}}$$

Remarquez que si les deux variables sont non corrélées : $Corr(x,y) = Cov(x,y) = 0$, cette corrélation devient :

$$Corr(x+y, x-y) = \frac{V(x)-V(y)}{V(x)+V(y)} \quad \textbf{si} \ Corr(x,y)=0$$

34. Précisez la notion d'exogénéité stricte (forte) ? Soit un modèle :

$$\begin{cases} \boldsymbol{y_t = \alpha + \beta x_t + \varepsilon_t} \\ \boldsymbol{x_t = \delta + \gamma y_{t-1} + \nu_t} \end{cases} \quad \boldsymbol{t = 1, 2, \dots, T.}$$

avec $\boldsymbol{\varepsilon}$ et $\boldsymbol{\nu}$ indépendant. Est-ce que $\boldsymbol{x_t}$ est strictement exogène dans la première équation ? Quelles sont les conséquences sur les propriétés de l'estimateur des MCO ?

La réponse nécessite la connaissance du Chapitre 3 du cours !

La notion de stricte exogénéité veut dire qu'il n'y a pas de corrélation entre un terme d'erreur d'une observation et toutes les variables explicatives, non seulement de l'observation mais aussi de toutes les autres observations.

$$E(\varepsilon_t|\boldsymbol{X}) = 0 \quad \Leftrightarrow \quad E(\varepsilon_t|\boldsymbol{x}_1, \boldsymbol{x}_2, \dots \boldsymbol{x}_T) = 0$$

Si on remplace $\boldsymbol{x}_t$ dans la première équation, on obtient :

$$\begin{aligned} y_t &= \alpha + \beta(\delta + \gamma y_{t-1} + \nu_t)x_t + \varepsilon_t \\ &= (\alpha + \beta\delta) + \beta\gamma y_{t-1} + (\beta\nu_t + \varepsilon_t) \\ &= \theta_1 + \theta_2 y_{t-1} + \eta_t \quad \text{avec} \begin{cases} \theta_1 = \alpha + \beta\delta \\ \theta_2 = \beta\gamma \\ \eta_t = \beta\nu_t + \varepsilon_t \end{cases} \end{aligned}$$

Dans cette régression, on voit immédiatement que l'erreur au temps t est corrélée avec la variable explicative au temps $t+1$: y_{t+1} du fait de la spécification du modèle $y_t = \alpha + \beta x_t + \varepsilon_t$. En conséquence, l'estimateur MCO du modèle est biaisé en échantillon fini.

Cependant l'estimateur MCO est convergent. Le biais asymptotique tend vers zéro lorsque la taille de l'échantillon tend vers l'infini. Pour cela il faut montrer que :

$$\text{plim}\big(\widehat{\theta_2} - \theta_2\big) = \text{plim}\frac{\frac{1}{T-1}\sum_{t=2}^{T}(y_{t-1}-\bar{y})(\eta_t - \bar{\eta})}{\frac{1}{T-1}\sum_{t=2}^{T}(y_{t-1}-\bar{y})^2}$$

Attention les moyennes $\bar{y}$ et $\bar{\eta}$ sont calculées sur les $T-1$ dernières observations, mais asymptotiquement cela ne change rien…

Du fait du Lemme 2.3, la limite en probabilité d'un ratio est le ratio des limites en probabilité.

$$\text{plim}\left(\widehat{\theta_2}-\theta_2\right)=\frac{\text{plim}\left(\frac{1}{T-1}\sum_{t=2}^{T}(y_{t-1}-\bar{y})(\eta_t-\bar{\eta})\right)}{\text{plim}\left(\frac{1}{T-1}\sum_{t=2}^{T}(y_{t-1}-\bar{y})^2\right)}$$

Pour un processus stationnaire ergodique, on a le théorème ergodique sur les moyennes :

$$\text{plim}\left(\frac{1}{T-1}\sum_{t=2}^{T}(y_{t-1}-\bar{y})(\eta_t-\bar{\eta})\right)=Cov(y_{t-1},\eta_t)$$

$$\text{plim}\left(\frac{1}{T-1}\sum_{t=2}^{T}(y_{t-1}-\bar{y})^2\right)=V(y_{t-1})$$

Mais la covariance entre y_{t-1} et η_t est nulle.

$$Cov(y_{t-1},\eta_t)=Cov(y_{t-1},\beta v_t+\varepsilon_t)=\beta\,Cov(y_{t-1},v_t)+Cov(y_{t-1},\varepsilon_t)=0$$

parce que v_t et ε_t sont indépendants et sont des innovations par rapport à y_{t-1}, ce dernier est une somme des erreurs v_t et ε_t passées. Donc le processus de y_t est une martingale, stationnaire, ergodique.

En conséquence, $\text{plim}\left(\widehat{\theta_2}-\theta_2\right)=0$, et l'estimateur MCO est convergent.

35. Pourquoi utilise-t-on $N-K$ pour le dénominateur de l'estimateur des MCO de la variance : $\widehat{\sigma^2}=SCR/(N-K)$? Montrez que cet estimateur est sans biais ? Quelle est la distribution de cet estimateur sous l'hypothèse de normalité des erreurs ?

[Voir démonstration dans les notes de cours]

36. Démontrez que la trace de la matrice idempotente $\boldsymbol{M}$ est égale à $\boldsymbol{N-K}$.

Cet exercice sert à démontrer que $E(SCR) = E(\boldsymbol{\varepsilon}'\boldsymbol{M}\boldsymbol{\varepsilon}) = \sigma^2(N-K)$ sous les hypothèse H1 à H4. En effet : $E(\boldsymbol{\varepsilon}'\boldsymbol{M}\boldsymbol{\varepsilon}) = E[tr(\boldsymbol{\varepsilon}'\boldsymbol{M}\boldsymbol{\varepsilon})]$ parce que la SCR est un scalaire et que la trace d'un scalaire est ce scalaire. Maintenant avec la propriété de circularité de la trace, on a : $E(\boldsymbol{\varepsilon}'\boldsymbol{M}\boldsymbol{\varepsilon}) = E[tr(\boldsymbol{M}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')]$. Les opérateurs d'espérance et de trace sont des opérateurs linéaires. Donc on peut les intervertir, et la matrice $\boldsymbol{M}$ dépend des variables explicatives :

$$E(\boldsymbol{\varepsilon}'\boldsymbol{M}\boldsymbol{\varepsilon}|\boldsymbol{X}) = tr[E(\boldsymbol{M}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|\boldsymbol{X})] = tr[\boldsymbol{M}E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|\boldsymbol{X})] = tr[\boldsymbol{M}\sigma^2\boldsymbol{I}_N]$$

par les hypothèse H4a (homoscédasticité) et H4b (absence d'autocorrélation) : $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|\boldsymbol{X}) = \sigma^2\boldsymbol{I}_N$. Ce qui donne $E(SCR|\boldsymbol{X}) = \sigma^2 tr(\boldsymbol{M})$ du fait des propriétés de la trace.

Par définition : $\boldsymbol{M} = \boldsymbol{I}_N - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ est une matrice $N \times N$ symétrique, idempotente et singulière avec la matrice des variables explicatives $\boldsymbol{X}$ de dimension $N \times N$.

$$tr(\boldsymbol{M}) = tr(\boldsymbol{I}_N - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')$$

L'opérateur de trace est un opérateur linéaire (c'est une somme). Dès lors, la trace d'une somme est la somme des traces :

$$tr(\boldsymbol{M}) = tr(\boldsymbol{I}_N) - tr(\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')$$

On utilise aussi la propriété de circularité de la trace pour obtenir :

$$tr(\boldsymbol{M}) = tr(\boldsymbol{I}_N) - tr((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}) = tr(\boldsymbol{I}_N) - tr(\boldsymbol{I}_K)$$

parce que la matrice carrée $\boldsymbol{X}'\boldsymbol{X}$ est de dimension $K \times K$. On utilise ici la définition de l'inverse d'une matrice. La trace de la matrice identité est évidemment l'ordre de cette matrice : $tr(\boldsymbol{I}_N) = N$ et $tr(\boldsymbol{I}_K) = K$. Ce qui donne finalement :

$$tr(\boldsymbol{M}) = N - K$$

37. Soit un estimateur sans biais $\widetilde{\boldsymbol{\beta}}$ (ce n'est pas nécessairement l'estimateur des MCO $\widehat{\boldsymbol{\beta}}$),

a) Prouvez que $\boldsymbol{Var}(\widetilde{\boldsymbol{\beta}}) = \boldsymbol{E}[\boldsymbol{Var}(\widetilde{\boldsymbol{\beta}}|\boldsymbol{X})] + \boldsymbol{Var}[\boldsymbol{E}(\widetilde{\boldsymbol{\beta}}|\boldsymbol{X})]$.

b) Prouvez que $\boldsymbol{Var}(\widetilde{\boldsymbol{\beta}}) \geq \boldsymbol{Var}(\widehat{\boldsymbol{\beta}})$.

Indices : Par définition,

$$Var(\tilde{\boldsymbol{\beta}}|\boldsymbol{X}) = E\left[\left(\tilde{\boldsymbol{\beta}} - E(\tilde{\boldsymbol{\beta}}|\boldsymbol{X})\right)\left(\tilde{\boldsymbol{\beta}} - E(\tilde{\boldsymbol{\beta}}|\boldsymbol{X})\right)'\middle|\boldsymbol{X}\right]$$

$$Var[E(\tilde{\boldsymbol{\beta}}|\boldsymbol{X})] = E\left\{\left(E(\tilde{\boldsymbol{\beta}}|\boldsymbol{X}) - E(\tilde{\boldsymbol{\beta}})\right)\left(E(\tilde{\boldsymbol{\beta}}|\boldsymbol{X}) - E(\tilde{\boldsymbol{\beta}})\right)'\right\}$$

Si $Var(\tilde{\boldsymbol{\beta}}|\boldsymbol{X}) \geq Var(\hat{\boldsymbol{\beta}}|\boldsymbol{X})$, alors $E[Var(\tilde{\boldsymbol{\beta}}|\boldsymbol{X})] \geq E[Var(\hat{\boldsymbol{\beta}}|\boldsymbol{X})]$

[Hayashi (2000), exercice 4, Section I.3]

A. Définissons

$$\boldsymbol{d} = \tilde{\boldsymbol{\beta}} - E(\tilde{\boldsymbol{\beta}}|\boldsymbol{X})$$
$$\boldsymbol{a} = \tilde{\boldsymbol{\beta}} - E(\tilde{\boldsymbol{\beta}})$$
$$\boldsymbol{c} = E(\tilde{\boldsymbol{\beta}}|\boldsymbol{X}) - E(\tilde{\boldsymbol{\beta}})$$

On aura alors : $\boldsymbol{d} = \boldsymbol{a} - \boldsymbol{c}$ et $\boldsymbol{dd'} = \boldsymbol{aa'} - \boldsymbol{ca'} - \boldsymbol{ac'} + \boldsymbol{cc'}$. On prend alors l'espérance <u>inconditionnelle</u> de cette expression :

$$E(\boldsymbol{dd'}) = E(\boldsymbol{aa'}) - E(\boldsymbol{ca'}) - E(\boldsymbol{ac'}) + E(\boldsymbol{cc'}).$$

Maintenant par la loi des espérances totales :

$$\begin{aligned} E(\boldsymbol{dd'}) &= E\{E[\boldsymbol{dd'}|\boldsymbol{X}]\} \\ &= E\left\{E\left[\left(\tilde{\boldsymbol{\beta}} - E(\tilde{\boldsymbol{\beta}}|\boldsymbol{X})\right)\left(\tilde{\boldsymbol{\beta}} - E(\tilde{\boldsymbol{\beta}}|\boldsymbol{X})\right)'\middle|\boldsymbol{X}\right]\right\} \\ &= E\{Var(\tilde{\boldsymbol{\beta}}|\boldsymbol{X})\} \end{aligned}$$

par la définition de la variance conditionnelle.

De même :

$$E(\boldsymbol{aa'}) = E\left[\left(\tilde{\boldsymbol{\beta}} - E(\tilde{\boldsymbol{\beta}})\right)\left(\tilde{\boldsymbol{\beta}} - E(\tilde{\boldsymbol{\beta}})\right)'\right] = Var(\tilde{\boldsymbol{\beta}})$$

$$\begin{aligned} E(\boldsymbol{cc'}) &= E\left[\left(E(\tilde{\boldsymbol{\beta}}|\boldsymbol{X}) - E(\tilde{\boldsymbol{\beta}})\right)\left(E(\tilde{\boldsymbol{\beta}}|\boldsymbol{X}) - E(\tilde{\boldsymbol{\beta}})\right)'\right] \\ &= E\left[\left(E(\tilde{\boldsymbol{\beta}}|\boldsymbol{X}) - E(\tilde{\boldsymbol{\beta}})\right)\left(E(\tilde{\boldsymbol{\beta}}|\boldsymbol{X}) - E(\tilde{\boldsymbol{\beta}})\right)'\right] = Var[E(\tilde{\boldsymbol{\beta}}|\boldsymbol{X})] \end{aligned}$$

par la définition de la variance et parce que $E_{\boldsymbol{X}}[E(\tilde{\boldsymbol{\beta}}|\boldsymbol{X})] = E(\tilde{\boldsymbol{\beta}})$.

$$\begin{aligned} E(\boldsymbol{ca'}) &= E\{E[\boldsymbol{ca'}|\boldsymbol{X}]\} \\ &= E\left\{E\left[\left(E(\tilde{\boldsymbol{\beta}}|\boldsymbol{X}) - E(\tilde{\boldsymbol{\beta}})\right)\left(\tilde{\boldsymbol{\beta}} - E(\tilde{\boldsymbol{\beta}})\right)'\middle|\boldsymbol{X}\right]\right\} \end{aligned}$$

Conditionnellement à $\boldsymbol{X}$, $E(\tilde{\boldsymbol{\beta}}|\boldsymbol{X}) - E(\tilde{\boldsymbol{\beta}})$ est un vecteur de variables fixes. On obtient alors :

$$E(\boldsymbol{ca}') = E\left\{\left(E(\tilde{\boldsymbol{\beta}}|\boldsymbol{X}) - E(\tilde{\boldsymbol{\beta}})\right) \times E\left[\left(\tilde{\boldsymbol{\beta}} - E(\tilde{\boldsymbol{\beta}})\right)\Big|\boldsymbol{X}\right]'\right\}$$

Mais $E_{\boldsymbol{X}}\left[\left(\tilde{\boldsymbol{\beta}} - E(\tilde{\boldsymbol{\beta}})\right)\Big|\boldsymbol{X}\right] = E_{\boldsymbol{X}}(\tilde{\boldsymbol{\beta}}|\boldsymbol{X}) - E(\tilde{\boldsymbol{\beta}})$. Comme $E_{\boldsymbol{X}}[E(\tilde{\boldsymbol{\beta}}|\boldsymbol{X})] = E(\tilde{\boldsymbol{\beta}})$, on aura donc :

$$E(\boldsymbol{ca}') = E\left\{\left(E(\tilde{\boldsymbol{\beta}}|\boldsymbol{X}) - E(\tilde{\boldsymbol{\beta}})\right)\left(E(\tilde{\boldsymbol{\beta}}|\boldsymbol{X}) - E(\tilde{\boldsymbol{\beta}})\right)'\right\} = Var\{E(\tilde{\boldsymbol{\beta}}|\boldsymbol{X})\}$$

En conséquence :

$$E(\boldsymbol{dd}') = E(\boldsymbol{aa}') - E(\boldsymbol{ca}') - E(\boldsymbol{ac}') + E(\boldsymbol{cc}').$$

$$\begin{aligned} E\{Var(\tilde{\boldsymbol{\beta}}|\boldsymbol{X})\} &= Var(\tilde{\boldsymbol{\beta}}) - Var[E(\tilde{\boldsymbol{\beta}}|\boldsymbol{X})] - Var[E(\tilde{\boldsymbol{\beta}}|\boldsymbol{X})] + Var[E(\tilde{\boldsymbol{\beta}}|\boldsymbol{X})] \\ &= Var(\tilde{\boldsymbol{\beta}}) - Var[E(\tilde{\boldsymbol{\beta}}|\boldsymbol{X})] \end{aligned}$$

Donc cela prouve la première question A :

$$V(\tilde{\boldsymbol{\beta}}) = E\{Var(\tilde{\boldsymbol{\beta}}|\boldsymbol{X})\} + Var[E(\tilde{\boldsymbol{\beta}}|\boldsymbol{X})]$$

B. Si l'estimateur $\tilde{\boldsymbol{\beta}}$ est sans biais : $E(\tilde{\boldsymbol{\beta}}|\boldsymbol{X}) = \boldsymbol{\beta}$. En conséquence : $Var[E(\tilde{\boldsymbol{\beta}}|\boldsymbol{X})] = Var(\boldsymbol{\beta}) = \mathbf{0}$ parce qu'on calcule la variance d'un vecteur de paramètre fixe. Le résultat précédent nous donne alors : $Var(\tilde{\boldsymbol{\beta}}) = E[Var(\tilde{\boldsymbol{\beta}}|\boldsymbol{X})]$

De même pour l'estimateur des *MCO* qui est sans biais : $E(\hat{\boldsymbol{\beta}}|\boldsymbol{X}) = \boldsymbol{\beta}$ et $Var[E(\hat{\boldsymbol{\beta}}|\boldsymbol{X})] = Var(\boldsymbol{\beta}) = \mathbf{0}$, ce qui donne : $Var(\hat{\boldsymbol{\beta}}) = E[Var(\hat{\boldsymbol{\beta}}|\boldsymbol{X})]$

Comme indiqué dans les indices,

$$\text{Si } Var(\tilde{\boldsymbol{\beta}}|\boldsymbol{X}) \geq Var(\hat{\boldsymbol{\beta}}|\boldsymbol{X}), \text{ alors } \quad E[Var(\tilde{\boldsymbol{\beta}}|\boldsymbol{X})] \geq E[Var(\hat{\boldsymbol{\beta}}|\boldsymbol{X})]$$

En conséquence :

$$Var(\tilde{\boldsymbol{\beta}}) \geq Var(\hat{\boldsymbol{\beta}})$$

Ce qui est la version inconditionnelle du théorème de Gauss-Markov.

38. Prouvez la propriété II.1.i :

$$\boldsymbol{Cov}\left(\boldsymbol{e},\widehat{\boldsymbol{\beta}}\middle|\boldsymbol{X}\right)=\boldsymbol{0}$$

[Hayashi (2000), exercice 6, Section I.3]

39. Démontrez que la variance d'un paramètre estimé par MCO peut s'écrire :

$$V\left(\widehat{\beta_k}\middle|\boldsymbol{X}\right)=\frac{\sigma^2}{SCT_k\left(1-R_{(k)}^2\right)}$$

avec $SCT_k=\sum_{i=1}^{N}\left(x_{k,i}-\overline{x_k}\right)^2$ la somme des carrés totaux de la $k^{ème}$ variable explicative, et $R_{(k)}^2$: le R^2 de la régression de x_k sur toutes les autres variables explicatives.

Les équations normales de l'estimation par MCO implique pour la $k^{ème}$ variable explicative :

$$\sum_{i=1}^{N}x_{i,k}\left(y_i-\widehat{\beta_1}-\widehat{\beta_2}x_{i,2}-\cdots-\widehat{\beta_k}x_{i,k}-\cdots-\widehat{\beta_K}x_{i,K}\right)=0$$

Réécrivons la $k^{ème}$ variable explicative $x_{i,k}$ comme la valeur calculée $\widehat{x_{i,k}}$ d'une régression de cette variable explicatives sur toutes les autres variables plus le résidus $r_{i,k}$ de cette régression : $x_{i,k}=\widehat{x_{i,k}}+r_{i,k}$. On insère alors ce résultat dans l'expression précédente :

$$\sum_{i=1}^{N}\left(\widehat{x_{i,k}}+r_{i,k}\right)\left(y_i-\widehat{\beta_1}-\widehat{\beta_2}x_{i,2}-\cdots-\widehat{\beta_k}x_{i,k}-\cdots-\widehat{\beta_K}x_{i,K}\right)=0$$

Le résidus e_i de cette estimation par MCO du modèle avec toutes les variables explicatives est orthogonal à la $k^{ème}$ variable explicative $x_{i,k}$: $\sum_{i=1}^{N}x_{i,k}e_i=0$. Mais $\widehat{x_{i,k}}$ est seulement une combinaison linéaire des autres variables explicatives : $x_{i,1},x_{i,2},\ldots,x_{i,K}$. En conséquence $\sum_{i=1}^{N}\widehat{x_{i,k}}e_i=0$. Ce qui donne :

$$\sum_{i=1}^{N}r_{i,k}\left(y_i-\widehat{\beta_1}-\widehat{\beta_2}x_{i,2}-\cdots-\widehat{\beta_k}x_{i,k}-\cdots-\widehat{\beta_K}x_{i,K}\right)=0$$

Mais $r_{i,k}$ est le résidus de la régression de $x_{i,k}$ sur toutes les autres variables explicatives et :

$$\sum_{i=1}^{N} r_{i,k}x_{i,j} = 0 \quad \text{pour tout } j = 1,2,\ldots,K \text{ et } j \neq k$$

Cela simplifie l'expression précédente :

$$\sum_{i=1}^{N} r_{i,k}\left(y_i - \widehat{\beta_k}x_{i,k}\right) = 0$$

$$\sum_{i=1}^{N} r_{i,k}\left(y_i - \widehat{\beta_k}\left(\widehat{x_{i,k}} + r_{i,k}\right)\right) = \sum_{i=1}^{N} r_{i,k}\left(y_i - \widehat{\beta_k}r_{i,k}\right) = 0$$

parce que $\sum_{i=1}^{N} r_{i,k}\widehat{\beta_k}\widehat{x_{i,k}} = \widehat{\beta_k}\sum_{i=1}^{N} r_{i,k}\widehat{x_{i,k}} = 0$ du fait que le résidus$r_{i,k}$ est orthogonal à la variable explicative calculée de cette régression $\widehat{x_{i,k}}$. L'estimateur $\widehat{\beta_k}$ dans la régression complète est enfin la solution de cette équation :

$$\sum_{i=1}^{N} r_{i,k}y_i - \widehat{\beta_k}\sum_{i=1}^{N} r_{i,k}^2 = 0 \quad \rightarrow \quad \widehat{\beta_k} = \frac{\sum_{i=1}^{N} r_{i,k}y_i}{\sum_{i=1}^{N} r_{i,k}^2}$$

Si on remplace la variable dépendante par le vrai modèle : $y_i = \beta_1 + \beta_2 x_{i,2} + \cdots + \beta_k x_{i,k} + \cdots + \beta_K x_{i,K} + \varepsilon_i$, et tenant compte que $\sum_{i=1}^{N} r_{i,k}x_{i,j} = 0$ pour $j \neq k$, on aura :

$$\widehat{\beta_k} = \frac{\sum_{i=1}^{N} r_{i,k}\left(\beta_k x_{i,k} + \varepsilon_i\right)}{\sum_{i=1}^{N} r_{i,k}^2}$$

Mais $\sum_{i=1}^{N} r_{i,k}x_{i,k} = \sum_{i=1}^{N} r_{i,k}\left(\widehat{x_{i,k}} + r_{i,k}\right) = \sum_{i=1}^{N} r_{i,k}^2$, cela donne finalement pour $\widehat{\beta_k}$:

$$\widehat{\beta_k} = \beta_k + \frac{\sum_{i=1}^{N} r_{i,k}\varepsilon_i}{\sum_{i=1}^{N} r_{i,k}^2}$$

Il est immédiat que cet estimateur est sans biais : $E\left(\widehat{\beta_k}\middle|\boldsymbol{X}\right) = \beta_k$ si l'hypothèse H2 est vérifiée.

La variance de l'estimateur des MCO de β_k est alors :

$$Var\left(\widehat{\beta_k}\middle|\boldsymbol{X}\right) = Var\left(\beta_k + \frac{\sum_{i=1}^{N} r_{i,k}\varepsilon_i}{\sum_{i=1}^{N} r_{i,k}^2}\middle|\boldsymbol{X}\right) = Var\left(\frac{\sum_{i=1}^{N} r_{i,k}\varepsilon_i}{\sum_{i=1}^{N} r_{i,k}^2}\middle|\boldsymbol{X}\right) = \frac{\sum_{i=1}^{N} r_{i,k}^2 Var(\varepsilon_i|\boldsymbol{X})}{\left(\sum_{i=1}^{N} r_{i,k}^2\right)^2}$$

si on suppose qu'il n'y a pas de covariance entre les erreurs d'observations différentes (hypothèse H4b) : $Cov\left(\varepsilon_i, \varepsilon_j\right) = 0$ pour $i \neq j$. Finalement, la variance des erreurs est identique pour toutes les observations : $Var(\varepsilon_i) = \sigma^2$ pour tout i (Hypothèse H4a d'homoscédasticité). En conséquence :

$$Var\left(\widehat{\beta_k}\middle|\boldsymbol{X}\right) = \frac{\sum_{i=1}^{N} r_{i,k}^2\,\sigma^2}{\left(\sum_{i=1}^{N} r_{i,k}^2\right)^2} = \frac{\sigma^2}{\sum_{i=1}^{N} r_{i,k}^2}$$

Le dénominateur est la somme des résidus de la régression de la variable x_k sur les autres variables explicatives du modèle : $SCR_k = \sum_{i=1}^{N} r_{i,k}^2$. Si on note R_k^2 le coefficient de détermination de cette régression et $SCT_k = \sum_{i=1}^{N} x_{i,k}^2$ la somme des carrés de la variable x_k , on aura par définition : $R_k^2 = 1 - (SCR_k/SCT_k)$. Ce qui donne : $SCR_k = SCT_k(1 - R_k^2)$.

On obtient finalement :

$$Var(\widehat{\beta_k}|\boldsymbol{X}) = \frac{\sigma^2}{SCT_k(1 - R_k^2)}$$

40. Supposons que le vrai modèle de régression multiple soit le suivant :

$$y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \varepsilon$$

Les hypothèses H1 à H5 sont satisfaites. Cependant on estime le modèle qui omet la variable x_4. Soient $\widetilde{\beta_1}, \widetilde{\beta_2}$, et $\widetilde{\beta_3}$ les estimateurs MCO de cette régression de y sur x_2 et x_3. Montrez que l'espérance de $\widetilde{\beta_2}$ (étant donné les valeurs des variables explicatives dans l'échantillon est :

$$E\left(\widetilde{\beta}_2\right) = \beta_2 + \beta_4 \frac{\sum_{i=1}^{N} u_{i2} x_{i4}}{\sum_{i=1}^{N} u_{i2}^2}$$

$$E\left(\widetilde{\beta_2}\right) = \beta_2 + \beta_4 \frac{\sum_{i=1}^{N} u_{2,i} x_{4,i}}{\sum_{i=1}^{N} {u_{2,i}}^2}$$

avec $u_{2,i}$ les résidus de la régression par MCO de x_2 sur x_3 ?

[Wooldridge (2009), exercice 3.10]

41. Considérez le modèle de régression simple : $y = \beta_1 + \beta_2 x + \varepsilon$. Sous les hypothèses de Gauss-Markov, les estimateurs traditionnels des MCO $\widehat{\beta_1}$ et $\widehat{\beta_2}$ sont des estimateurs sans biais des paramètres de la population. Soit $\widetilde{\beta_2}$ l'estimateur de lorsqu'il n'y pas de constante dans la régression.

a) Exprimer $E(\widetilde{\beta_2})$ en fonction des x_i, β_1 et β_2. Vérifiez que $\widetilde{\beta_2}$ est un estimateur sans biais de β_2 lorsque la valeur de la constante au sein de la population (β_1) est nulle. Existe-t-il d'autres cas pour lesquels $\widetilde{\beta_2}$ est sans biais ?

L'estimateur des MCO du modèle $y = \beta_2 x + \varepsilon$ sans la constante est par définition :

$$\widetilde{\beta_2} = \frac{\sum_{i=1}^{N} x_i y_i}{\sum_{i=1}^{N} x_i^2}$$

On remplace alors la variable dépendante par le vrai modèle (incluant la constante) :

$$\widetilde{\beta_2} = \frac{\sum_{i=1}^{N} x_i(\beta_1 + \beta_2 x_i + \varepsilon_i)}{\sum_{i=1}^{N} x_i^2} = \beta_1 \frac{\sum_{i=1}^{N} x_i}{\sum_{i=1}^{N} x_i^2} + \beta_2 + \frac{\sum_{i=1}^{N} x_i \varepsilon_i}{\sum_{i=1}^{N} x_i^2}$$

Calculons alors l'espérance de $\widetilde{\beta_2}$ (conditionnelle à $\boldsymbol{x}$) en supposant l'hypothèse H2 (les erreurs ne sont corrélés avec les variables explicatives $E(\varepsilon_i|\boldsymbol{X}) = 0$:

$$E(\widetilde{\beta_2}) = \beta_2 + \beta_1 \frac{\sum_{i=1}^{N} x_i}{\sum_{i=1}^{N} x_i^2}$$

On voit immédiatement que l'estimateur $\widetilde{\beta_2}$ sera sans biais si la vraie valeur de la constante est nulle : $\beta_1 = 0$. De même, si la moyenne de la variable explicative $\bar{x}$ est nulle, il n'y aura également pas de biais pour $\widetilde{\beta_2}$.

b) Calculez la variance de $\widetilde{\beta_2}$.

La variance de $\widetilde{\beta_2}$ est par définition :

$$V(\widetilde{\beta_2}) = E\left(\widetilde{\beta_2} - E(\widetilde{\beta_2})\right)^2$$

$$= E\left(\beta_2 + \beta_1 \frac{\sum_{i=1}^{N} x_i}{\sum_{i=1}^{N} x_i^2} + \frac{\sum_{i=1}^{N} x_i \varepsilon_i}{\sum_{i=1}^{N} x_i^2} - \beta_2 - \beta_1 \frac{\sum_{i=1}^{N} x_i}{\sum_{i=1}^{N} x_i^2}\right)^2$$

$$= E\left(\frac{\sum_{i=1}^{N} x_i \varepsilon_i}{\sum_{i=1}^{N} x_i^2}\right)^2$$

Comme la variance est calculée conditionnellement aux variables explicatives $\boldsymbol{x}$:

$$V(\widetilde{\beta_2}) = \frac{\sum_{i=1}^{N} x_i^2 \sigma^2}{(\sum_{i=1}^{N} x_i^2)^2} = \frac{\sigma^2}{\sum_{i=1}^{N} x_i^2}$$

sous les hypothèse H4a (homoscédasticité) et H4b (absence d'autocorrélation).

c) **Montrez que $V(\widetilde{\beta_2}) \leq V(\widehat{\beta_2})$.**

[<u>Astuce</u> : pour tout échantillon : $\sum_i x_i^2 \geq \sum_i (x_i - \bar{x})^2$ avec une inégalité stricte, sauf lorsque $\bar{x} = 0$.]

Dans une régression simple avec constante, on a vu que :

$$\widehat{\beta_2} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{N}(x_i - \bar{x})^2} \quad et \quad V(\widehat{\beta_2}) = \frac{\sigma^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

Mais $\sum_{i=1}^{N}(x_i - \bar{x})^2 = \sum_{i=1}^{N} x_i^2 - N\bar{x}^2 \leq \sum_{i=1}^{N} x_i^2$ (avec l'égalité si la moyenne $\bar{x}$ est nulle.

Donc :

$$V(\widetilde{\beta_2}) = \frac{\sigma^2}{\sum_{i=1}^{N} x_i^2} \leq \frac{\sigma^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2} = V(\widehat{\beta_2})$$

L'estimateur $\widetilde{\beta_2}$ de β_2 est en général biaisé (sauf si $\beta_1 = 0$ ou si $\bar{x} = 0$), mais il est plus précis parce que sa variance est plus faible !

d) **Expliquez le compromis qui existe entre le biais et la variance lorsqu'il s'agit de faire un choix entre $\widehat{\beta_2}$ et $\widetilde{\beta_2}$.**

Une manière de comparer un estimateur sans biais, avec un estimateur biaisé mais plus précis est d'utiliser le concept d'erreur quadratique moyenne (Mean-Square Error) défini comme :

$$MSE = E(\hat{\theta} - \theta)^2$$

(attention on soustrait la vraie valeur du paramètre, et non pas l'espérance de l'estimateur $E(\hat{\theta})$). Cela peut se réécrire comme :

$$MSE(\hat{\theta}) = E(\hat{\theta} - E(\hat{\theta}) + E(\hat{\theta}) - \theta)^2$$

Remarquez que par définition : $Biais(\hat{\theta}) = E(\hat{\theta}) - \theta$ qui est une valeur fixe. Donc :

$$MSE(\hat{\theta}) = E\left(\hat{\theta} - E(\hat{\theta}) + Biais(\hat{\theta})\right)^2$$

$$= E\left(\hat{\theta} - E(\hat{\theta})\right)^2 + 2E\left[\left(\hat{\theta} - E(\hat{\theta})\right) Biais(\hat{\theta})\right] + E[Biais(\hat{\theta})]^2$$

Remarquez aussi que, par définition, le premier terme est la variance de l'estimateur : $E\left(\hat{\theta} - E(\hat{\theta})\right)^2 = Var(\hat{\theta})$. Donc :

$$MSE(\hat{\theta}) = Var(\hat{\theta}) + 2Biais(\hat{\theta})E\left(\hat{\theta} - E(\hat{\theta})\right) + Biais(\hat{\theta})^2$$

$$= Var(\hat{\theta}) + Biais(\hat{\theta})^2$$

Du fait que $E\left(\hat{\theta} - E(\hat{\theta})\right) = E(\hat{\theta}) - E(\hat{\theta}) = 0$.

Calculons l'erreur quadratique moyenne pour les deux estimateurs, sachant que l'estimateur des MCO du modèle avec constante est sans biais :

$$MSE(\widehat{\beta_2}) = Var(\widehat{\beta_2}) = \frac{\sigma^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

$$MSE(\widetilde{\beta_2}) = Var(\widetilde{\beta_2}) + Biais(\widetilde{\beta_2})^2 = \frac{\sigma^2}{\sum_{i=1}^{N} x_i^2} + \left(\beta_1 \frac{\sum_{i=1}^{N} x_i}{\sum_{i=1}^{N} x_i^2}\right)^2$$

$$= \frac{\sigma^2 + \beta_1^2 \frac{(\sum_{i=1}^{N} x_i)^2}{\sum_{i=1}^{N} x_i^2}}{\sum_{i=1}^{N} x_i^2}$$

Le numérateur et le dénominateur étant plus grands que dans l'erreur quadratique moyenne de $\widehat{\beta_2}$, il est difficile de comparer ces deux erreurs quadratiques moyennes, car cela dépend aussi de la valeur de β_1 !

42. Supposons que les variables $y_i, x_{1,i},$ et $x_{2,i}$ satisfont les hypothèses des moindres carrés et que, pour toutes les observations $V(\varepsilon_i | x_{1,i}, x_{2,i}) = 4$ et $V(x_{1,i}) = 6$. L'échantillon aléatoire comprend 400 observations.

a) Supposons que x_1 et x_2 soient non corrélés. Calculez la variance de $\hat{\beta}_1$? Pour cela démontrez que :

$$V(\widehat{\beta_1}) = \frac{1}{N} \frac{1}{1 - (\rho_{x_1,x_2})^2} \frac{{\sigma_\varepsilon}^2}{{\sigma_{x_1}}^2}$$

où ρ_{x_1,x_2} est la corrélation théorique entre x_1 et x_2, ${\sigma_{x_1}}^2$ est la variance théorique de x_1.

b) Supposons maintenant que la corrélation entre x_1 et x_2 soit égale à 0.5 : $\rho_{x_1,x_2} = 0.5$. Calculez la variance de $\widehat{\beta_1}$?

c) **Commentez l'affirmation : « *si* x_1 et x_2 *sont corrélés, la variance de* $\widehat{\beta_1}$ *est plus grande que dans le cas contraire* » ?**
[Stock et Watson (2011), exercice 3.8]

43. Considérons le modèle de régression sans constante :

$$y_i = \beta_1 x_{1,i} + \beta_2 x_{2,i} + \varepsilon_i, \qquad i = 1,2, \dots, N.$$

Répondez aux questions suivantes :

a) **Précisez la fonction des moindres carrés minimisée par les MCO ?**

b) **Calculez les conditions du premier ordre et donnez les équations normales ?**

c) **Supposons que** $\sum_{i=1}^{N} x_{1,i}x_{2,i} = 0$, **montrez qu'on peut alors écrire :**

$$\widehat{\beta_1} = \frac{\sum_{i=1}^{N} x_{1,i}y_i}{\sum_{i=1}^{N} {x_{1,i}}^2} \quad et \quad \widehat{\beta_2} = \frac{\sum_{i=1}^{N} x_{2,i}y_i}{\sum_{i=1}^{N} {x_{2,i}}^2}.$$

d) **Supposons que** $\sum_{i=1}^{N} x_{1,i}x_{2,i} \neq 0$, **établissez une expression de** $\widehat{\beta_1}$ **en fonction des variables** $y_i, x_{1,i},$ et $x_{2,i}$?

e) **Supposons que le modèle comporte une constante, et s'écrit donc :** $y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \varepsilon_i$. **Démontrez que les estimateurs des moindres carrés satisfont :** $\widehat{\beta_0} = \bar{y} - \widehat{\beta_1}\overline{x_1} - \widehat{\beta_2}\overline{x_2}$?

f) **De la même manière que dans (e), supposons que le modèle comporte une constante et que** $\sum_{i=1}^{N}\left(x_{1,i} - \overline{x_1}\right)\left(x_{2,i} - \overline{x_2}\right) = 0$. **Montrez que**

$$\widehat{\beta_1} = \frac{\sum_{i=1}^{N}\left(x_{1,i} - \overline{x_1}\right)(y_i - \bar{y})}{\sum_{i=1}^{N}\left(x_{1,i} - \overline{x_1}\right)^2}$$

g) **Que se passe-t-il si la variable x_2 est omise dans la régression ? Comparez les estimateurs de β_1 des MCO correspondants aux régressions avec et sans la variable x_2 ?**
[Stock et Watson (2011), exercice 3.9]

44. Dans une régression simple, $\widehat{\beta_1}$ et $\widehat{\beta_2}$ sont les estimateurs de l'ordonnée à l'origine et de la pente de la droite de régression. Soit $\bar{\varepsilon}$ la moyenne des erreurs (et non des résidus !).

a) **Montrez que $\widehat{\beta_2}$ peut s'écrire sous la forme $\widehat{\beta_2} = \beta_2 + \sum_{i=1}^{N} w_i \varepsilon_i$ où $w_i = d_i / SCT$ et $d_i = x_i - \bar{x}$.**

b) **En partant du point (i), montrez que $\sum_i w_i = 0$ et que $\widehat{\beta_2}$ et $\bar{\varepsilon}$ ne sont pas corrélés.**

c) **Montrez que $\widehat{\beta_1}$ peut s'écrire sous la forme : $\widehat{\beta_1} = \beta_1 + \bar{\varepsilon} - \left(\widehat{\beta_1} - \beta_1\right)\bar{x}$.**

d) **Utilisez les points (ii) et (iii) pour montrer que $V\left(\widehat{\beta_1}\right) = \frac{\sigma^2}{N} + \frac{\sigma^2 \bar{x}^2}{SCT_x}$.**

e) **Simplifiez l'expression précédente pour aboutir à l'équation :**

$$V\left(\widehat{\beta_1}\right) = \frac{\sigma^2}{N} \frac{\sum_i x_i^2}{\sum_i (x_i - \bar{x})^2}$$

45. Supposons que nous avons deux estimateurs $\left(\widehat{\theta_1} \text{ et } \widehat{\theta_2}\right)$ d'un paramètre θ.Ceux-ci sont sans biais, et indépendants l'un de l'autre. Leur variance est respectivement v_1 et v_2. Quelle est la combinaison linéaire des deux estimateurs : $\widehat{\theta} = c_1 \widehat{\theta_1} + c_2 \widehat{\theta_2}$ qui est l'estimateur sans biais de variance minimale de θ ?
[Greene (2018), exercice 4.1]

Une combinaison linéaire d'estimateur sans biais sera aussi sans biais :

$$E(\widehat{\boldsymbol{\theta}}) = E\left[c_1\widehat{\boldsymbol{\theta_1}} + c_2\widehat{\boldsymbol{\theta_2}}\right] = c_1E\left[\widehat{\boldsymbol{\theta_1}}\right] + c_2E\left[\widehat{\boldsymbol{\theta_2}}\right] = c_1\boldsymbol{\theta} + c_2\boldsymbol{\theta} = \boldsymbol{\theta}$$

si et seulement si $c_1 + c_2 = 1$, ou $c_2 = 1 - c_1$.

Calculons la variance de cette combinaison linéaire :

$$V(\widehat{\boldsymbol{\theta}}) = V\left[c_1\widehat{\boldsymbol{\theta_1}} + c_2\widehat{\boldsymbol{\theta_2}}\right] = c_1^2V\left[\widehat{\boldsymbol{\theta_1}}\right] + c_2^2V\left[\widehat{\boldsymbol{\theta_2}}\right] + 2c_1c_2Cov\left(\widehat{\boldsymbol{\theta_1}}, \widehat{\boldsymbol{\theta_2}}\right)$$

Mais les deux estimateurs sont indépendants : $Cov\left(\widehat{\boldsymbol{\theta_1}}, \widehat{\boldsymbol{\theta_2}}\right) = 0$, et $c_2 = 1 - c_1$:

$$V(\widehat{\boldsymbol{\theta}}) = c_1^2v_1 + (1 - c_1)^2v_2$$

La combinaison linéaire des estimateurs qui minimise cette variance sera :

$$c_1^* = \arg\min_{c_1} V(\widehat{\boldsymbol{\theta}}) = c_1^2v_1 + (1 - c_1)^2v_2$$

La condition du premier ordre est :

$$\frac{\partial V(\widehat{\boldsymbol{\theta}})}{\partial c_1} = 2c_1^*v_1 - 2(1 - c_1^*)v_2 = 0$$

que l'on résous facilement :

$$2c_1^*v_1 + 2c_1^*v_2 = 2v_2$$

$$c_1^* = \frac{v_2}{v_1 + v_2}$$

et

$$c_2^* = 1 - c_1^* = \frac{v_1}{v_1 + v_2}$$

Les poids de la combinaison linéaire des deux estimateurs qui minimise la variance est proportionnelle à la part respective de la variance de l'autre estimateur sur la variance totale de la somme des deux estimateurs.

46. Supposons un modèle de régression linéaire classique, mais que la vraie valeur de la constante est zéro.
Comparez la variance de l'estimateur MCO des paramètres de pente calculé sans la constante avec celle de l'estimateur MCO des paramètres de pente calculé avec cette constante inutile.

[Greene (2018), exercice 4.3]

[Voir exercice 41]

47. Considérons la régression multiple de y sur K variables explicatives X et une variable explicative additionnelle z.
Sous les hypothèses H1 à H4 du modèle de régression linéaire classique, prouvez que la vraie variance de l'estimateur des MCO des paramètres de X est plus grande quand la variable z est incluse dans la régression que dans le cas où elle n'est incluse dans la régression. (on suppose que le coefficient de z n'est pas zéro)
[Greene (2018), exercice 4.8]

Soit le modèle complet : $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{z}\gamma + \boldsymbol{\varepsilon}$. Du fait du Théorème de Frisch – Waugh, l'estimateur des MCO de $\boldsymbol{\beta}$ sera :

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{M_z X})^{-1}\boldsymbol{X}'\boldsymbol{M_z y} \quad \text{avec } \boldsymbol{M_z} = \boldsymbol{I_N} - \boldsymbol{z}(\boldsymbol{z}'\boldsymbol{z})^{-\mathbf{1}}\boldsymbol{z}'$$

Dont la matrice de variance-covariance est donné par :

$$V(\widehat{\boldsymbol{\beta}}) = \sigma^2(\boldsymbol{X}'\boldsymbol{M_z X})^{-1}$$

Si on effectue la régression « courte » sans la variable explicative $\boldsymbol{z}$ du modèle $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\eta}$, on aura l'estimateur des MCO : $\overline{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$ avec une matrice de variance-covariance :

$$V(\overline{\boldsymbol{\beta}}) = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

Mais $\boldsymbol{X}'\boldsymbol{X} - \boldsymbol{X}'\boldsymbol{M_z X} = \boldsymbol{X}'\boldsymbol{X} - \boldsymbol{X}'(\boldsymbol{I_N} - \boldsymbol{z}(\boldsymbol{z}'\boldsymbol{z})^{-\mathbf{1}}\boldsymbol{z}')\boldsymbol{X} = \boldsymbol{X}'\boldsymbol{z}(\boldsymbol{z}'\boldsymbol{z})^{-\mathbf{1}}\boldsymbol{z}'\boldsymbol{X} > \mathbf{0}$, une matrice définie positive. La matrice $\boldsymbol{X}'\boldsymbol{X}$ est plus grande (au sens matriciel) que la matrice $\boldsymbol{X}'\boldsymbol{M_z X}$, ce qui implique que $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ est plus petite (au sens matriciel) que $(\boldsymbol{X}'\boldsymbol{M_z X})^{-1}$. Dès lors :

$$V(\overline{\boldsymbol{\beta}}) = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1} < \sigma^2(\boldsymbol{X}'\boldsymbol{M_z X})^{-1} = V(\widehat{\boldsymbol{\beta}})$$

L'estimateur $\overline{\boldsymbol{\beta}}$ des MCO de $\boldsymbol{\beta}$ dans la régression sans la variable explicative $\boldsymbol{z}$ aura une variance plus faible que l'estimateur $\widehat{\boldsymbol{\beta}}$ de $\boldsymbol{\beta}$ dans la régression avec la variable explicative $\boldsymbol{z}$. $\overline{\boldsymbol{\beta}}$ **sera un estimateur plus précis que** $\widehat{\boldsymbol{\beta}}$, mais il peut être biaisé si la variable $\boldsymbol{z}$ joue dans l'explication de la variable dépendante.

Attention ce raisonnement suppose que l'on connaisse la vraie valeur de la variance de l'erreur σ^2. Dans la pratique, il faut estimer cette variance. On peut le faire classiquement par :

$$\widehat{\sigma^2} = \frac{SCR_L}{N-K}$$

pour la régression « longue » avec les K variables explicatives $\boldsymbol{X}$ et $\boldsymbol{z}$; et avec :

$$\overline{\sigma^2} = \frac{SCR_C}{N-(K-1)}$$

pour la régression « courte » sans la variable explicative $\boldsymbol{z}$. On sait que le numérateur $SCR_L \leq SCR_C$, mais que le dénominateur est aussi plus petit dans la régression longue : $N-K < N-K+1$. On ne peut affirmer laquelle de ces deux estimations est la plus grande. Mais en général, $\widehat{\sigma^2} \leq \overline{\sigma^2}$, ce qui va dans le sens inverse de la conclusion précédente pour la comparaison des variances des paramètres estimés !

48. Un chercheur dispose de deux échantillons indépendants des variables (Y_i, X_i) qui désignent respectivement les salaires et le nombre d'années d'étude pour les hommes et pour les femmes.
Pour les hommes, (1er échantillon), considérons la régression :

$$Y_{M,i} = \beta_{1,M} + \beta_{2,M} X_{M,i} + \varepsilon_{M,i}$$

Pour les femmes, (2ème échantillon), considérons la régression :

$$Y_{F,i} = \beta_{1,F} + \beta_{2,F} X_{M,i} + \varepsilon_{F,i}$$

$s_{\widehat{\beta_{2,M}}}$ et $s_{\widehat{\beta_{2,F}}}$ représentent respectivement l'écart-type estimé du paramètre de la variable X dans les deux régressions. Montrez que l'écart-type estimé de $\widehat{\beta_{2,M}} - \widehat{\beta_{2,F}}$ s'écrit :

$$s_{(\widehat{\beta_{2,M}} - \widehat{\beta_{2,F}})} = \sqrt{\left(s_{\widehat{\beta_{2,M}}}\right)^2 + \left(s_{\widehat{\beta_{2,F}}}\right)^2}$$

[Stock et Watson (2011), exercice 2.15]

49. Si les contraintes sont correctes, démontrez l'estimateur des moindres carrés contraint est sans biais. Quelles sont les hypothèses que vous devez supposer ? Montrez que l'estimateur des MCO est aussi sans biais.
Dérivez la matrice de variance – covariance de l'estimateur des MCC. Montrez qu'elle est plus petite (au sens matriciel) que celle des MCO. Interprétez ce résultat.

50. Donnez une expression du biais de l'estimateur des MCC si la contrainte n'est pas correcte ?

51. On sait que la somme des carrés des résidus des MCC est supérieure à celle des MCO. Donnez la condition pour que l'estimateur de la variance de l'erreur pour les MCC (σ^{*2}) soit supérieure à l'estimateur de la variance de l'erreur des MCO ($\hat{\sigma}^2$).

52. Considérons le modèle de régression multiple avec 3 variables explicatives sous les hypothèses usuelles du modèle linéaire classique :

$$\boldsymbol{y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon}$$

Vous souhaitez tester l'hypothèse nulle H_0: $\beta_1 - 3\beta_2 = 1$.

a) Soit $\widehat{\beta_1}$ et $\widehat{\beta_2}$ les estimateurs des MCO des paramètres β_1 et β_2 respectivement. Ecrivez $Var(\widehat{\beta_1} - 3\widehat{\beta_2})$ en fonction des variances et covariances des paramètres estimés $\widehat{\beta_1}$ et $\widehat{\beta_2}$. Quelle est l'expression de l'écart-type de $\widehat{\beta_1} - 3\widehat{\beta_2}$?

Variance de $\widehat{\beta_1} - 3\widehat{\beta_2}$:

$$\begin{aligned} V(\widehat{\beta_1} - 3\widehat{\beta_2}) &= V(\widehat{\beta_1}) + V(-3\widehat{\beta_2}) + 2 \times Cov(\widehat{\beta_1}, -3\widehat{\beta_2}) \\ &= V(\widehat{\beta_1}) + 9V(\widehat{\beta_2}) - 6Cov(\widehat{\beta_1}, \widehat{\beta_2}) \end{aligned}$$

Ecart-type :

$$s(\widehat{\beta_1} - 3\widehat{\beta_2}) = \sqrt{V(\widehat{\beta_1}) + 9V(\widehat{\beta_2}) - 6Cov(\widehat{\beta_1}, \widehat{\beta_2})}$$

b) Ecrivez la statistique t permettant de tester H_0: $\beta_1 - 3\beta_2 = 1$.

$$t_{\beta_1 - 3\beta_2 - 1 = 0} = \frac{\widehat{\beta_1} - 3\widehat{\beta_2} - 1}{s(\widehat{\beta_1} - 3\widehat{\beta_2})}$$

c) On définit $\theta_1 = \beta_1 - 3\beta_2$. Réécrivez l'équation de départ de sorte à faire apparaître les paramètres $\beta_0, \theta_1, \beta_2$ et β_3 dans l'équation et ainsi obtenir directement l'estimateur de $\widehat{\theta_1}$.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

On va ajouter et soustraire $3\beta_2 x_1$, ce qui ne change rien à ce modèle :

$$y = \beta_0 + \beta_1 x_1 - 3\beta_2 x_1 + 3\beta_2 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

On regroupe les termes en x_1 :

$$y = \beta_0 + (\beta_1 - 3\beta_2) x_1 + \beta_2 (3x_1 + x_2) + \beta_3 x_3 + \varepsilon$$

En notant $z = 3x_1 + x_2$, une nouvelle variable de ce modèle :

$$y = \beta_0 + \theta_1 x_1 + \beta_2 z + \beta_3 x_3 + \varepsilon$$

La régression de y sur une constante, x_1, z et x_3 concerne le même modèle mais qui est ici reparamétrisé avec les paramètres $\beta_0, \theta_1, \beta_2$ et β_3.

On peut alors tester directement l'hypothèse nulle H_0: $\theta_1 = 1$ avec un test t sur le second paramètre du modèle.

d) On définit aussi $\theta_2 = \beta_3 + 2\beta_2$. Quels sont les variances de $\widehat{\theta_1}$ et $\widehat{\theta_2}$, et la covariance entre $\widehat{\theta_1}$ et $\widehat{\theta_2}$ en fonction des variances et covariances des $\widehat{\beta_j}$.

$$V(\widehat{\theta_2}) = V(\widehat{\beta_3} + 2\widehat{\beta_2}) = V(\widehat{\beta_3}) + 4V(\widehat{\beta_2}) + 4Cov(\widehat{\beta_3}, \widehat{\beta_2})$$

$$Cov(\widehat{\theta_1}, \widehat{\theta_2}) = Cov[(\widehat{\beta_1} - 3\widehat{\beta_2}), (\widehat{\beta_3} + 2\widehat{\beta_2})]$$

$$= Cov(\widehat{\beta_1}, \widehat{\beta_3}) - 3Cov(\widehat{\beta_2}, \widehat{\beta_3}) + 2Cov(\widehat{\beta_1}, \widehat{\beta_2}) - 6V(\widehat{\beta_2})$$

53. Pour un nombre d'observation (N) et un nombre de paramètres donnés (K), quel est le niveau du R^2 qui assure le rejet de l'hypothèse nulle que tous les paramètres de pente soient zéros ? Utilisez pour cela le test F de significativité globale d'une régression. Quel est ce niveau de R^2 pour $K = 5$ et $N = 20, 50, 100$ ou 1000 ?

On utilise le test F de significativité globale des paramètres de pente d'une régression qui s'écrit en fonction des R^2 du modèle (Chapiter II - slide 98) :

$$F = \frac{R^2}{1 - R^2} \times \frac{N - K}{K - 1}$$

Si la statistique F est supérieure au seuil critique $F_{1-\alpha}(K - 1, N - K)$ pour un niveau de test de $\alpha\%$, on rejette l'hypothèse nulle et il y a au moins un paramètre non nul dans le modèle. Considérons le R_*^2 limite qui correspond à ce seuil critique :

$$F = \frac{R_*^2}{1 - R_*^2} \times \frac{N - K}{K - 1} \geq F_{1-\alpha}(K - 1, N - K)$$

Pour la clarté de l'écriture, on ne va plus indiquer les degrés de liberté de cette loi F de Fisher. La condition devient :

$$\frac{R_*^2}{1 - R_*^2} \geq \frac{K - 1}{N - K} F_{1-\alpha}$$

$$\frac{N - K}{K - 1} \frac{1}{F_{1-\alpha}} \geq \frac{1}{R_*^2} - 1$$

$$\frac{N - K}{K - 1} \frac{1}{F_{1-\alpha}} + 1 \geq \frac{1}{R_*^2}$$

$$R_*^2 \geq \frac{1}{1 + \frac{N - K}{K - 1} \frac{1}{F_{1-\alpha}}} = \frac{F_{1-\alpha}}{F_{1-\alpha} + \frac{N - K}{K - 1}}$$

Si le coefficient de détermination de la régression est supérieur ou égal à ce seuil, on rejettera l'hypothèse nulle que tous les paramètres de pente sont zéros !

Pour un niveau de test de 5% ($\alpha = 5\%$), et pour 5 variables explicatives dans le modèle (y compris la constante), on teste la nullité conjointe des 4 paramètres de pente. On recherche les seuils critiques dans la table de la loi F de Fisher :

$$\begin{cases} pour\ N = 20 & : \quad F_{0.95}(4{,}15) = 3.056 \\ pour\ N = 50 & : \quad F_{0.95}(4{,}45) = 2.579 \\ pour\ N = 100 & : \quad F_{0.95}(4{,}95) = 2.467 \\ pour\ N = 1000 & : \quad F_{0.95}(4{,}995) = 2.381 \end{cases}$$

[On peut utiliser la fonction EXCEL : `INVERSE.LOI.F(`α ; $K - 1$; $N - K$`)`].

On peut remarquer que lorsque $N \to \infty$, ce seuil critique : $F_{1-\alpha}(K-1, N-K) \xrightarrow{d} \frac{\chi^2_{1-\alpha}(K-1)}{K-1}$. On aura ici une convergence vers : $\chi^2_{0.95}(4)/4 = 9.488/4 = 2.372$!

On effectue alors le calcul pour $N = 20$:

$$R_*^2 \geq \frac{3.056}{3.056 + \frac{15}{4}} = \frac{3.056}{3.056 + \frac{15}{4}} = \frac{3.056}{3.056 + 3.75} = 0.449$$

De même, pour $N = 50$, $R_*^2 \geq 0.186$. pour $N = 100$, $R_*^2 \geq 0.094$ et pour $N = 1000$, $R_*^2 \geq 0.009$. Plus la taille de l'échantillon augmente, plus le R_*^2 limite est proche de zéro,. Cela indique qu'on peut se contenter d'un coefficient de détermination faible pour avoir une significativité globale de la régression lorsque la taille de l'échantillon est élevée !

ÉCONOMETRIE THEORIQUE

(M1 MBFA)

Corrigés des Exercices Théoriques

Chapitre 3 : Les propriétés asymptotiques de l'estimateur des MCO
Quelques corrections

54. **Supposons que $\sqrt{N}\left(\widehat{\theta_N}-\theta\right) \xrightarrow{d} N(0\,,\sigma^2)$. Prouvez que $\widehat{\theta_N} \xrightarrow{p} \theta$.**
[Hayashi, 2000, Section II.1, Exercice 4]

On peut réécrire l'hypothèse : $\left(\widehat{\theta_N}-\theta\right)=\frac{1}{\sqrt{N}}\sqrt{N}\left(\widehat{\theta_N}-\theta\right)$. Or $\lim\limits_{N\to\infty}\frac{1}{\sqrt{N}}=\operatorname{plim}\limits_{N\to\infty}\frac{1}{\sqrt{N}}=0$. Cela donne avec la partie b) du Lemme 2.4 :

$$\left(\widehat{\theta_N}-\theta\right)=\underbrace{\frac{1}{\sqrt{N}}}_{\xrightarrow{p}0}\underbrace{\sqrt{N}\left(\widehat{\theta_N}-\theta\right)}_{\xrightarrow{d}N(0,\sigma^2)}\xrightarrow{p}0$$

En conséquence : $\widehat{\theta_N}\xrightarrow{p}\theta$.

55. **Soit $\{z_i\}$ une séquence de variables aléatoires indépendantes et identiquement distribuées avec $E(z_i)=\mu\neq 0$ et $V(z_i)=\sigma^2$. La moyenne d'échantillon est $\overline{z_N}=\frac{1}{N}\sum_{i=1}^{N}z_i$. Démontrez que**

$$\sqrt{N}\left(\frac{1}{\overline{z_N}}-\frac{1}{\mu}\right)\xrightarrow{d}N\left(0,\frac{\sigma^2}{\mu^4}\right)$$

[Hayashi, 2000, Section II.1, Exercice 5]

56. **Soit un vecteur $K \times 1$ de variables aléatoires z_t stationnaires en covariance avec une matrice des autocovariances $\Gamma_s = Cov(z_t, z_{t-s})$. Montrez que $\Gamma_s = \Gamma'_{-s}$**
[Hayashi, 2000, Section II.2, Exercice 1]

L'autocovariance d'ordre s pour la vecteur $\boldsymbol{z}_t$ est définie par :

$$Cov(\boldsymbol{z}_t, \boldsymbol{z}_{t-s}) = E[(\boldsymbol{z}_t - \boldsymbol{\mu})(\boldsymbol{z}_{t-s} - \boldsymbol{\mu})'] = \boldsymbol{\Gamma}_s$$

avec $\boldsymbol{\mu} = E(\boldsymbol{z}_t)$. Si $\{\boldsymbol{z}_t\}$ est stationnaire en covariance, on sait que :

$$\begin{aligned}\boldsymbol{\Gamma}_s = Cov(\boldsymbol{z}_t, \boldsymbol{z}_{t-s}) &= Cov(\boldsymbol{z}_{t+s}, \boldsymbol{z}_t) \\ &= E[(\boldsymbol{z}_{t+s} - \boldsymbol{\mu})(\boldsymbol{z}_t - \boldsymbol{\mu})'] \\ &= E\left[\left((\boldsymbol{z}_t - \boldsymbol{\mu})(\boldsymbol{z}_{t+s} - \boldsymbol{\mu})'\right)\right]' = \boldsymbol{\Gamma}'_{-s}\end{aligned}$$

57. **Supposons que le processus $\{x_t\}$ soit une martingale par rapport à $\{z_t\}$. Montrez que $E(x_{t+s}|z_{t-1}, z_{t-2}, \dots z_1) = x_{t-1}$ et que $E(x_{t+s+1} - x_{t+s}|z_{t-1}, z_{t-2}, \dots z_1) = 0$ pour $s = 0, 1, \dots$**
[Hayashi, 2000, Section II.2, Exercice 3]

Si le processus $\{x_t\}$ soit une martingale par rapport à $\{z_t\}$, $E(x_t|z_{t-1}, z_{t-1}, \dots z_1) = x_{t-1}$.

$$E(x_{t+s}|z_{t+s-1}, z_{t+s-2}, \dots z_1) = x_{t+s-1}$$

Mais

$$E(x_{t+s-1}|z_{t+s-2}, z_{t+s-3}, \dots z_1) = x_{t+s-2}$$

Donc :

$$\begin{aligned}E[E(x_{t+s}|z_{t+s-1}, z_{t+s-2}, \dots z_1)|z_{t+s-2}, z_{t+s-3}, \dots z_1] &= \\ E(x_{t+s}|z_{t+s-2}, z_{t+s-3}, \dots z_1) &= x_{t+s-2}\end{aligned}$$

En appliquant ce raisonnement récursivement, on aura :

$$E(x_{t+s}|z_{t-1}, z_{t-2}, \dots z_1) = x_{t-1}$$

Maintenant : $E(x_{t+s}|z_{t-1}, z_{t-2}, \dots z_1) = x_{t-1}$ et $E(x_{t+s+1}|z_{t-1}, z_{t-2}, \dots z_1) = x_{t-1}$, donc on aura immédiatement :

$$\begin{aligned}E(x_{t+s+1} - x_{t+s}|z_{t-1}, z_{t-2}, \dots z_1) &= E(x_{t+s+1}|z_{t-1}, z_{t-2}, \dots z_1) - E(x_{t+s}|z_{t-1}, z_{t-2}, \dots z_1) \\ &= x_{t-1} - x_{t-1} = 0.\end{aligned}$$

58. **Soit $\{x_i\}$ une séquence de nombres réels et $\{\varepsilon_i\}$ une séquence de variables aléatoires indépendantes et identiquement distribuées d'espérance nulle et de variance finie. Est-ce que la séquence $\{x_i\varepsilon_i\}$ est indépendante et identiquement distribuée ? A-t-elle des autocorrélations nulles ? Est-elle une martingale en différence ? Est-elle stationnaire ?**
[Hayashi, 2000, Section II.2, Exercice 4]

59. **Définissez une marche aléatoire. Quelle est son espérance, sa variance et ses autocorrélations. Montrez que cette marche aléatoire n'est pas stationnaire.**
[Hayashi, 2000, Section II.2, Exercice 5]

Par la définition 16, une marche aléatoire $\{z_T\}$ est le cumul d'un processus bruit blanc $\{x_T\}$:

$$z_t = \sum_{s=1}^{t} x_s \qquad \text{avec } x_s \approx i.i.d.(0, \sigma^2)$$

L'espérance de la marche aléatoire est :

$$E(z_t) = E\left(\sum_{s=1}^{t} x_s\right) = \sum_{s=1}^{t} \underbrace{E(x_s)}_{=0} = 0.$$

La variance de la marche aléatoire est :

$$V(z_t) = V\left(\sum_{s=1}^{t} x_s\right) = \sum_{s=1}^{t} \underbrace{V(x_s)}_{=\sigma^2} + 2\sum_{s=1}^{t}\sum_{r=s+1}^{t} \underbrace{Cov(x_s, x_r)}_{=0} = \sum_{s=1}^{t} \sigma^2 = t \times \sigma^2$$

parce que les x_s sont des variables aléatoires homoscédastiques et indépendantes (bruit blanc). Remarquez que la marche aléatoire a une variance qui augmente au cours du temps. Elle est donc hétéroscédastique ! Elle ne peut pas être stationnaire.

La covariance d'ordre r entre deux éléments de cette marche aléatoire est alors :

$$Cov(z_t, z_{t-r}) = E[z_t z_{t-r}] = E\left[\sum_{s=1}^{t} x_s \sum_{s=1}^{t-r} x_s\right] = \sum_{s=1}^{t-r} E(x_s^2) = (t-r)\sigma^2$$

L'autocorrélation d'ordre r entre deux éléments de cette marche aléatoire dépendent de l'écart de temps r, mais aussi du moment de l'observation t :

$$Corr(z_t, z_{t-r}) = \frac{Cov(z_t, z_{t-r})}{\sqrt{V(z_t) \times V(z_{t-r})}} = \frac{(t-r)\sigma^2}{\sqrt{t \times \sigma^2 \times (t-r)\sigma^2}}$$

$$= \frac{(t-r)}{\sqrt{t \times (t-r)}} = \frac{\sqrt{t-r}}{\sqrt{t}} = \sqrt{1 - \frac{r}{t}}$$

Les autocorrélations pour une marche aléatoire tendent asymptotiquement vers 1 lorsque la taille de l'échantillon tend vers l'infini (quel que soit l'ordre de l'autocorrélation). Mais pour un échantillon fini, les autocorrélations décroissent avec leur ordre.

60. **Supposons le modèle de régression sans variables explicatives : $\boldsymbol{y_i = \mu + \varepsilon_i}$ avec $\boldsymbol{\varepsilon_i \approx i.i.d. N(0, \sigma^2)}$. Prouvez que la moyenne d'échantillon $\boldsymbol{\hat{\mu} = \bar{y} = \frac{1}{N}\sum_i y_i}$ est un estimateur convergent et asymptotiquement normal. Maintenant considérez l'estimateur alternatif : $\boldsymbol{\tilde{\mu} = \sum_i w_i y_i}$ avec $\boldsymbol{w_i = i/\sum_i i}$. Montrez que cet estimateur est convergent et calculez se variance asymptotique. Pour rappel : $\boldsymbol{\sum_i i = n(n+1)/2}$ et $\boldsymbol{\sum_i i^2 = n(n+1)(2n+1)/6}$.** ***[Greene (2018), exercice 4.8]***

61. **Soit $\boldsymbol{s_{\widehat{\beta_k}}}$ l'écart-type estimé d'un paramètre estimé par MCO. Montrez que $\boldsymbol{s_{\widehat{\beta_k}} \xrightarrow{p} 0}$** ***[Hayashi, 2000, Section II.4, Exercice 1]***

On sait que la matrice de variance-covariance de l'estimateur MCO est, sous les hypothèse de Gauss-Markov :

$$V(\widehat{\boldsymbol{\beta}}|\boldsymbol{X}) = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1} = \frac{\sigma^2}{N}\left(\frac{\boldsymbol{X}'\boldsymbol{X}}{N}\right)^{-1} = \frac{\sigma^2}{N}\boldsymbol{S}_{\boldsymbol{xx}}^{-1}$$

Par l'hypothèse de stationnarité et d'ergodicité et le théorème ergodique : $\boldsymbol{S}_{\boldsymbol{xx}} \xrightarrow{p} \boldsymbol{\Sigma}_{\boldsymbol{xx}}$ qui est une matrice définie-positive. Le premier terme tend vers 0, lorsque la taille de l'échantillon tend vers l'infini : $\lim_{N\to\infty}(\sigma^2/N) = 0$. Dès lors par les conséquences du Lemme 2.3, on aura

$$V(\widehat{\boldsymbol{\beta}}|\boldsymbol{X}) = \frac{\sigma^2}{N}\boldsymbol{S}_{\boldsymbol{xx}}^{-1} \xrightarrow{\boldsymbol{p}} 0 \times \boldsymbol{\Sigma}_{\boldsymbol{xx}}^{-1} = 0$$

Donc l'écart-type d'un paramètre estimé tend aussi en probabilité vers 0 : $s_{\widehat{\beta_k}} \xrightarrow{p} 0$.

Autre démonstration : Reprenons la formule de la variance de l'estimateur d'un paramètre dans une régression multiple :

$$V(\widehat{\beta_k}|\boldsymbol{X}) = \frac{\sigma^2}{SCT_k\left(1 - R^2_{(k)}\right)}$$

avec : $SCT_k = \sum_{i=1}^{N}\left(x_{k,i} - \overline{x_k}\right)^2$ la somme des carrés totaux de la $k^{ème}$ variable explicative,

$R^2_{(k)}$: le R^2 de la régression de x_k sur toutes les autres variables explicatives.

σ^2 est un paramètre fixe (la variance de l'erreur). De même le $R^2_{(k)}$ est aussi fixe, car il dépend seulement de la corrélation multiple entre cette variable explicative et les autres. Mais SCT_k tend vers l'infini lorsque la taille de l'échantillon augmente indéfiniment (c'est une somme de carrés). Dès lors $V(\widehat{\beta_k}|\boldsymbol{X}) \xrightarrow{p} 0$ et donc l'écart-type (sa racine carrée) tend aussi en probabilité vers 0 : $s_{\widehat{\beta_k}} \xrightarrow{p} 0$.

62. **Supposons par simplicité que $K = 1$ et que $\widehat{\boldsymbol{\beta}}$ est l'estimateur MCO de $\boldsymbol{\beta}$. L'écart-type de $\widehat{\boldsymbol{\beta}}$ est $\sqrt{Avar(\widehat{\boldsymbol{\beta}})/N}$. On définit $\boldsymbol{\lambda} = -\log \boldsymbol{\beta}$ qui est estimé par $\hat{\boldsymbol{\lambda}} = -\log \widehat{\boldsymbol{\beta}}$. Vérifiez que l'écart-type de $\hat{\boldsymbol{\lambda}}$ est alors $(1/\widehat{\boldsymbol{\beta}})\sqrt{Avar(\widehat{\boldsymbol{\beta}})/N}$.**
[Hayashi, 2000, Section II.4, Exercice 2]

63. **Il n'y a pas une manière unique d'écrire les $\boldsymbol{J}$ hypothèses linéaires $\boldsymbol{R\beta} = \boldsymbol{r}$. On peut représenter les mêmes restrictions par $\widetilde{\boldsymbol{R}}\boldsymbol{\beta} = \tilde{\boldsymbol{r}}$ en prémultipliant $\boldsymbol{R}$ et $\boldsymbol{r}$ par une matrice $\boldsymbol{F}$ de dimension $\boldsymbol{J} \times \boldsymbol{J}$ non singulière : $\widetilde{\boldsymbol{R}} = \boldsymbol{FR}$ et $\tilde{\boldsymbol{r}} = \boldsymbol{Fr}$. Est-ce que les choix de $\boldsymbol{R}$ ou de $\widetilde{\boldsymbol{R}}$ modifient la valeur numérique du test de Wald, sa distribution asymptotique, sa distribution en échantillon fini ?**
[Hayashi, 2000, Section II.4, Exercice 3]

La statistique de test de Wald de l'hypothèse nulle H_0: $\boldsymbol{R\beta} = \boldsymbol{r}$ est :

$$W_1 = \left(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{r}\right)'\left(\boldsymbol{R}\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\beta}}}\boldsymbol{R}'\right)^{-1}\left(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{r}\right)$$

où $\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\beta}}}$ est la matrice de variance-covariance de l'estimateur $\widehat{\boldsymbol{\beta}}$.

La statistique de test de Wald de l'hypothèse nulle $H_0: \widetilde{\boldsymbol{R}}\boldsymbol{\beta} = \widetilde{\boldsymbol{r}}$ (avec $\widetilde{\boldsymbol{R}} = \boldsymbol{FR}$ et $\widetilde{\boldsymbol{r}} = \boldsymbol{Fr}$ est :

$$W_2 = \left(\widetilde{\boldsymbol{R}}\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{r}}\right)'\left(\widetilde{\boldsymbol{R}}\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\beta}}}\widetilde{\boldsymbol{R}}'\right)^{-1}\left(\widetilde{\boldsymbol{R}}\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{r}}\right)$$

$$= \left(\boldsymbol{FR}\widehat{\boldsymbol{\beta}} - \boldsymbol{Fr}\right)'\left(\boldsymbol{FR}\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\beta}}}(\boldsymbol{FR})'\right)^{-1}\left(\boldsymbol{FR}\widehat{\boldsymbol{\beta}} - \boldsymbol{Fr}\right)$$

$$= \left(\boldsymbol{F}\left(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{r}\right)\right)'\left(\boldsymbol{FR}\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\beta}}}(\boldsymbol{FR})'\right)^{-1}\left(\boldsymbol{F}\left(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{r}\right)\right)$$

$$= \left(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{r}\right)'\boldsymbol{F}'\left(\boldsymbol{FR}\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\beta}}}\boldsymbol{R}'\boldsymbol{F}'\right)^{-1}\boldsymbol{F}\left(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{r}\right)$$

Comme $\boldsymbol{F}$ est une matrice dimension $J \times J$ non singulière, son inverse existe et on aura :

$$W_2 = \left(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{r}\right)'\boldsymbol{F}'\boldsymbol{F}'^{-1}\left(\boldsymbol{R}\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\beta}}}\boldsymbol{R}'\right)^{-1}\boldsymbol{F}^{-1}\boldsymbol{F}\left(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{r}\right)$$

$$= \left(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{r}\right)'\left(\boldsymbol{R}\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\beta}}}\boldsymbol{R}'\right)^{-1}\left(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{r}\right) = W_1$$

Les deux statistiques de Wald sont numériquement identiques. Elles convergent toutes els deux vers une loi du Khi-deux avec J degrés de liberté, donc elles sont asymptotiquement équivalente. La distribution en échantillon fini est également la même.

64. **On veut tester la relation non linéaire entre les paramètres d'un modèle : $\boldsymbol{\beta_1\beta_2 = 1}$ avec le test de Wald. Montrez que la statistique de Wald de l'hypothèse nulle $\boldsymbol{H_0: \beta_1\beta_2 - 1 = 0}$ est mathématiquement différente de la statistique de test de l'hypothèse nulle $\boldsymbol{H_0: \beta_1 - (1/\beta_2) = 0}$. Est-ce que ces deux statistiques sont asymptotiquement équivalentes ?**

La statistique de test de restrictions non linéaires de l'hypothèse nulle $H_0: \boldsymbol{a}(\boldsymbol{\beta}) = 0$ est :

$$W = \boldsymbol{a}(\widehat{\boldsymbol{\beta}})'\left(\boldsymbol{A}(\widehat{\boldsymbol{\beta}})\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\beta}}}\boldsymbol{A}(\widehat{\boldsymbol{\beta}})'\right)^{-1}\boldsymbol{a}(\widehat{\boldsymbol{\beta}})$$

où $\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\beta}}}$ est la matrice de variance-covariance de l'estimateur $\widehat{\boldsymbol{\beta}}$ et $\boldsymbol{A}(\boldsymbol{\beta}) = \partial\boldsymbol{a}(\boldsymbol{\beta})/\partial\boldsymbol{\beta}'$ est la matrice $J \times K$ des dérivées premières de $\boldsymbol{a}(\boldsymbol{\beta})$. On a vu que, sous les hypothèses H2.1 à H2.5, $W \overset{d}{\to} \chi^2(J)$

Pour le test de l'hypothèse nulle $H_0: \beta_1\beta_2 - 1 = 0$, on a une seule restriction non linéaire ($J = 1$) avec la fonction $\boldsymbol{a}(\boldsymbol{\beta}) = \beta_1\beta_2 - 1$. Par soucis de simplicité, on ne considère que les deux paramètres $\widehat{\beta_1}$ et $\widehat{\beta_2}$, avec une matrice de variance-covariance :

$$\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\beta}}} = \begin{bmatrix} s_1^2 & s_{12} \\ s_{12} & s_2^2 \end{bmatrix}$$

Le vecteur 1×2 des dérivées premières est simplement : $\boldsymbol{A}(\boldsymbol{\beta}) = (\beta_2 \quad \beta_1)$ que l'on évalue avec les paramètres estimés. Le test de Wald devient :

$$W_1 = \left(\widehat{\beta_1}\widehat{\beta_2} - 1\right)' \left(\left(\widehat{\beta_2} \quad \widehat{\beta_1}\right) \begin{bmatrix} s_1^2 & s_{12} \\ s_{12} & s_2^2 \end{bmatrix} \begin{pmatrix} \widehat{\beta_2} \\ \widehat{\beta_1} \end{pmatrix} \right)^{-1} \left(\widehat{\beta_1}\widehat{\beta_2} - 1\right)$$

$$= \frac{\left(\widehat{\beta_1}\widehat{\beta_2} - 1\right)^2}{\widehat{\beta_2}^2 s_1^2 + 2\widehat{\beta_1}\widehat{\beta_2} s_{12} + \widehat{\beta_1}^2 s_2^2}$$

Sous l'hypothèse nulle, W_1 est asymptotiquement distribué selon une loi du Khi-deux avec 1 degrés de liberté.

Avec la formulation alternative de l'hypothèse nulle $H_0 \colon \beta_1 - (1/\beta_2) = 0$, on a encore une seule restriction non linéaire. Maintenant le vecteur des dérivées première est :

$$\boldsymbol{A}(\boldsymbol{\beta}) = \begin{pmatrix} 1 & \frac{1}{\beta_2^2} \end{pmatrix}$$

La variance asymptotique de $\widehat{\beta_1} - \left(1/\widehat{\beta_2}\right)$ est :

$$V\left(\widehat{\beta_1} - \left(1/\widehat{\beta_2}\right)\right) = \begin{pmatrix} 1 & \frac{1}{\beta_2^2} \end{pmatrix} \begin{bmatrix} s_1^2 & s_{12} \\ s_{12} & s_2^2 \end{bmatrix} \begin{pmatrix} 1 \\ 1/\beta_2^2 \end{pmatrix}$$

$$= \begin{pmatrix} s_1^2 + \frac{s_{12}}{\beta_2^2} & s_{12} + \frac{s_2^2}{\beta_2^2} \end{pmatrix} \begin{pmatrix} 1 \\ 1/\beta_2^2 \end{pmatrix}$$

$$= s_1^2 + 2\frac{s_{12}}{\beta_2^2} + \frac{s_2^2}{\beta_2^4}$$

La statistique de test de Wald non linéaire est alors :

$$W_2 = \left(\widehat{\beta_1} - \left(1/\widehat{\beta_2}\right)\right)' \left(s_1^2 + 2\frac{s_{12}}{\widehat{\beta_2^2}} + \frac{s_2^2}{\widehat{\beta_2^4}}\right)^{-1} \left(\widehat{\beta_1} - \left(1/\widehat{\beta_2}\right)\right) = \frac{\left(\widehat{\beta_1} - \left(1/\widehat{\beta_2}\right)\right)^2}{s_1^2 + 2\frac{s_{12}}{\widehat{\beta_2^2}} + \frac{s_2^2}{\widehat{\beta_2^4}}}$$

Sous l'hypothèse nulle, W_2 est aussi asymptotiquement distribué selon une loi du Khi-deux avec 1 degrés de liberté.

On peut réécrire cette statistique W_2 en multipliant par $1/\widehat{\beta_2^2}$:

$$W_2 = \frac{\frac{1}{\widehat{\beta_2^2}}\left(\widehat{\beta_1}\widehat{\beta_2} - 1\right)^2}{\frac{1}{\widehat{\beta_2^2}}\left(\widehat{\beta_2^2}s_1^2 + 2s_{12} + \frac{s_2^2}{\widehat{\beta_2^2}}\right)} = \frac{\left(\widehat{\beta_1}\widehat{\beta_2} - 1\right)^2}{\widehat{\beta_2^2}s_1^2 + 2s_{12} + \frac{s_2^2}{\widehat{\beta_2^2}}}$$

Le numérateur de W_1 et de W_2 sont identiques, mais pas les dénominateurs : pour le premier, on a : $\widehat{\beta_2^2}s_1^2 + 2\widehat{\beta_1}\widehat{\beta_2}s_{12} + \widehat{\beta_1^2}s_2^2$, alors que pour le second, on a : $\widehat{\beta_2^2}s_1^2 + 2s_{12} + \frac{s_2^2}{\widehat{\beta_2^2}}$, à moins d'avoir le cas exceptionnel : $\widehat{\beta_1} = 1/\widehat{\beta_2}$, c'est-à-dire si l'hypothèse nulle est parfaitement vérifiée.

65. **On veut tester la relation non linéaire entre les paramètres d'un modèle : $\boldsymbol{\beta_1\beta_2 = \beta_3}$ avec le test de Wald. Montrez que la statistique de Wald de l'hypothèse nulle $\boldsymbol{H_0: \beta_1\beta_2 - \beta_3 = 0}$ est mathématiquement différente de la statistique de test de l'hypothèse nulle $\boldsymbol{H_0: \beta_1 - (\beta_3/\beta_2) = 0}$. Est-ce que ces deux statistiques sont asymptotiquement équivalentes ?**

66. **Montrez que la vraie variance de l'estimateur des MCO est plus grande que la vraie variance de l'estimateur des MCG.**

67. **Parmi les propositions suivantes, quelles sont celles susceptibles d'être causées par la présence d'hétéroscédasticité ? Expliquez.**
    a) **Les estimateurs des MCO, $\boldsymbol{\widehat{\beta_k}}$, ne sont pas convergents.**
    b) **La statistique *F* habituelle ne suit plus une distribution *F* de Fisher.**
    c) **Les estimateurs des MCO ne sont plus les meilleurs estimateurs linéaires sans biais.**

68. **Avec les hypothèses 2.1 à 2.5, la variance asymptotique de l'estimateur des MCO est $Avar(\widehat{\beta}) = \Sigma_{xx}^{-1} S \Sigma_{xx}^{-1}$ avec $S = E(\varepsilon_i^2 x_i x_i')$. Est-elle estimée de manière convergente par $\widehat{Avar}(\widehat{\beta}) = N\widehat{\sigma^2} S_{xx}^{-1}$ avec $S_{xx} = \left(\frac{1}{N} X'X\right)^{-1}$ sans l'hypothèse (2.7) d'homoscédasticité conditionnelle ? Est-ce que la statistique de test $t_{\widehat{\beta_k}} = \widehat{\beta_k} / s_{\widehat{\beta_k}}$ est asymptotiquement distribué selon une loi normale standard.**
[Hayashi, 2000, Section II.6, Exercice 1]

69. **Sans l'hypothèse (2.7) d'homoscédasticité conditionnelle, est-ce que la statistique de test de J contraintes $(SCR_C - SCR)/\widehat{\sigma^2}$ est asymptotiquement distribué selon une loi du Khi-deux avec J degré de liberté.**
[Hayashi, 2000, Section II.6, Exercice 4]

70. **Dans une régression avec une constante, considérez l'hypothèse nulle que tous les $K-1$ paramètres de pente soient nuls. Montrez que $NR^2 \overset{d}{\rightarrow} \chi^2(K-1)$ sous les hypothèses 2.1 à 2.5 et l'hypothèse (2.7) d'homoscédasticité conditionnelle. Pourquoi ne peut-on pas utiliser cette expression lorsqu'il y a de l'hétéroscédasticité conditionnelle.**
[Hayashi, 2000, Section II.6, Exercice 5]

71. Quelle est la matrice de covariance, $Cov(\widehat{\beta_{MCG}}, \widehat{\beta_{MCG}} - \widehat{\beta_{MCO}})$, entre l'estimateur des MCG : $\widehat{\beta_{MCG}} = (X'\Psi^{-1}X)^{-1}X'\Psi^{-1}y$ et sa différence avec l'estimateur des MCO : $\widehat{\beta_{MCO}} = (X'X)^{-1}X'y$.
[Greene (2018), exercice 9.1]

L'estimateur MCO est :

$$\widehat{\beta_{MCO}} = (X'X)^{-1}X'y = \beta + (X'X)^{-1}X'\varepsilon$$

alors que l'estimateur MCG est :

$$\widehat{\beta_{MCG}} = (X'\Psi^{-1}X)^{-1}X'\Psi^{-1}y = \beta + (X'\Psi^{-1}X)^{-1}X'\Psi^{-1}\varepsilon$$

Leur différence peut s'écrire :

$$\widehat{\boldsymbol{\beta}_{MCG}} - \widehat{\boldsymbol{\beta}_{MCO}} = (\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{\varepsilon} - (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon}$$

$$= [(\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Psi}^{-1} - (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{\varepsilon}$$

Ces deux estimateurs sont sans biais, donc $E(\widehat{\boldsymbol{\beta}_{MCG}} - \widehat{\boldsymbol{\beta}_{MCO}}) = \mathbf{0}$. En conséquence la covariance entre $\widehat{\boldsymbol{\beta}_{MCG}}$ et la différence $\widehat{\boldsymbol{\beta}_{MCG}} - \widehat{\boldsymbol{\beta}_{MCO}}$ est :

$$Cov(\widehat{\boldsymbol{\beta}_{MCG}}, \widehat{\boldsymbol{\beta}_{MCG}} - \widehat{\boldsymbol{\beta}_{MCO}}) = E\left[(\widehat{\boldsymbol{\beta}_{MCG}} - \boldsymbol{\beta})(\widehat{\boldsymbol{\beta}_{MCG}} - \widehat{\boldsymbol{\beta}_{MCO}})'\right]$$

$$= E[(\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'((\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Psi}^{-1} - (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')']$$

$$= (\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Psi}^{-1}E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')[\boldsymbol{\Psi}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{X})^{-1} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}]$$

Mais comme $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \sigma^2\boldsymbol{\Psi}$, on obtient :

$$Cov(\widehat{\boldsymbol{\beta}_{MCG}}, \widehat{\boldsymbol{\beta}_{MCG}} - \widehat{\boldsymbol{\beta}_{MCO}}) = (\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\sigma^2\boldsymbol{\Psi}\boldsymbol{\Psi}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{X})^{-1}$$

$$-(\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\sigma^2\boldsymbol{\Psi}\,\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

$$= \sigma^2(\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{X})^{-1} - \sigma^2(\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\,\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

$$= \sigma^2(\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{X})^{-1} - \sigma^2(\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{X})^{-1} = \mathbf{0}$$

Cela permet de calculer la variance de la différence entre les deux estimateurs :

$$V(\widehat{\boldsymbol{\beta}_{MCG}} - \widehat{\boldsymbol{\beta}_{MCO}}) = V(\widehat{\boldsymbol{\beta}_{MCG}}) + V(\widehat{\boldsymbol{\beta}_{MCO}}) - 2Cov(\widehat{\boldsymbol{\beta}_{MCG}}, \widehat{\boldsymbol{\beta}_{MCO}})$$

Mais comme :

$$Cov(\widehat{\boldsymbol{\beta}_{MCG}}, \widehat{\boldsymbol{\beta}_{MCG}} - \widehat{\boldsymbol{\beta}_{MCO}}) = Cov(\widehat{\boldsymbol{\beta}_{MCG}}, \widehat{\boldsymbol{\beta}_{MCG}}) - Cov(\widehat{\boldsymbol{\beta}_{MCG}}, \widehat{\boldsymbol{\beta}_{MCO}}) = \mathbf{0}$$

$$V(\widehat{\boldsymbol{\beta}_{MCG}}) - Cov(\widehat{\boldsymbol{\beta}_{MCG}}, \widehat{\boldsymbol{\beta}_{MCO}}) = \mathbf{0}$$

$$Cov(\widehat{\boldsymbol{\beta}_{MCG}}, \widehat{\boldsymbol{\beta}_{MCO}}) = V(\widehat{\boldsymbol{\beta}_{MCG}})$$

Donc :

$$V(\widehat{\boldsymbol{\beta}_{MCG}} - \widehat{\boldsymbol{\beta}_{MCO}}) = V(\widehat{\boldsymbol{\beta}_{MCG}}) + V(\widehat{\boldsymbol{\beta}_{MCO}}) - 2V(\widehat{\boldsymbol{\beta}_{MCG}})$$

$$= V(\widehat{\boldsymbol{\beta}_{MCO}}) - V(\widehat{\boldsymbol{\beta}_{MCG}})$$

On sait d'ailleurs que $V(\widehat{\boldsymbol{\beta}_{MCO}}) > V(\widehat{\boldsymbol{\beta}_{MCG}})$, si les erreurs sont hétéroscédastiques et/ ou autocorrélées : $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \sigma^2\boldsymbol{\Psi}$

72. ***[Théorème de Kruskal]*** **Dans un modèle de régression par *MCG*, si on peut écrire que les K colonnes de X sont les vecteurs caractéristiques (*eigenvector*) de $\Omega = E(\varepsilon\varepsilon')$, alors les estimateurs des *MCG* et des *MCO* sont identiques.**
[Greene (2018), exercice 9.5]

73. **Dans un modèle de régression par *MCG* avec la matrice $\Omega = E(\varepsilon\varepsilon')$ connue,**

[Greene (2018), exercice 9.6]

a) Quelle est la matrice de covariance entre les estimateurs *MCG* ($\widehat{\boldsymbol{\beta}_{MCG}}$) et *MCO* ($\widehat{\boldsymbol{\beta}_{MCO}}$) ?

(Voir Exercice 71). Soit l'estimateur MCO : $\widehat{\boldsymbol{\beta}_{MCO}} = \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon}$ et l'estimateur MCG : $\widehat{\boldsymbol{\beta}_{MCG}} = \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{\varepsilon}$, la matrice de covariance entre les deux estimateurs est :

$$Cov\left(\widehat{\boldsymbol{\beta}_{MCO}}, \widehat{\boldsymbol{\beta}_{MCG}}\right) = E\left[\left(\widehat{\boldsymbol{\beta}_{MCO}} - \boldsymbol{\beta}\right)\left(\widehat{\boldsymbol{\beta}_{MCG}} - \boldsymbol{\beta}\right)'\right]$$

$$= E[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'((\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Psi}^{-1})']$$

$$= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}']\boldsymbol{\Psi}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{X})^{-1}$$

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

parce que $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \sigma^2\boldsymbol{\Psi}$. EN conséquence

$$Cov\left(\widehat{\boldsymbol{\beta}_{MCO}}, \widehat{\boldsymbol{\beta}_{MCG}}\right) = \sigma^2(\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{X})^{-1} = V\left(\widehat{\boldsymbol{\beta}_{MCG}}\right)$$

b) Quelle est la matrice de variance-covariance des résidus des *MCO* ($\boldsymbol{e}_{MCO} = \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}_{MCO}}$) ?

Le résidus MCO est : $\boldsymbol{e}_{MCO} = \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}_{MCO}} = \boldsymbol{y} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} = \boldsymbol{M}\boldsymbol{y} = \boldsymbol{M}\boldsymbol{\varepsilon}$. Sa matrice de variance – covariance (avec $E(\boldsymbol{\varepsilon}) = \boldsymbol{0}$, ce qui implique $E(\boldsymbol{e}_{MCO}) = \boldsymbol{0}$) est :

$$E(\boldsymbol{e}_{MCO}\boldsymbol{e}_{MCO}') = E(\boldsymbol{M}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{M}') = \boldsymbol{M}E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')\boldsymbol{M} = \sigma^2\boldsymbol{M}\boldsymbol{\Psi}\boldsymbol{M}$$

c) Quelle est la matrice de variance-covariance des résidus des *MCG* ($\boldsymbol{e}_{MCG} = \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}_{MCG}}$)

Le résidus MCG est : $\boldsymbol{e}_{MCG} = \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}_{MCG}} = \boldsymbol{y} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{y}$. Ce qui donne immédiatement :

$$\boldsymbol{e}_{MCG} = [\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Psi}^{-1}]\boldsymbol{\varepsilon}$$

Sa matrice de variance – covariance (avec $E(\boldsymbol{\varepsilon}) = \boldsymbol{0}$, ce qui implique $E(\boldsymbol{e}_{MCO}) = \boldsymbol{0}$) est :

$$\begin{aligned}
E(\boldsymbol{e}_{MCG}\boldsymbol{e}'_{MCG}) &= E[(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Psi}^{-1})\boldsymbol{\varepsilon\varepsilon}'(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Psi}^{-1})']\\
&= (\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Psi}^{-1})E[\boldsymbol{\varepsilon\varepsilon}'](\boldsymbol{I} - \boldsymbol{\Psi}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}')\\
&= \sigma^2(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Psi}^{-1})\boldsymbol{\Psi}(\boldsymbol{I} - \boldsymbol{\Psi}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}')\\
&= \sigma^2\boldsymbol{\Psi} - \sigma^2\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{\Psi} - \sigma^2\boldsymbol{\Psi}\boldsymbol{\Psi}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\\
&\qquad + \sigma^2\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{\Psi}\boldsymbol{\Psi}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\\
&= \sigma^2\boldsymbol{\Psi} - \sigma^2\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\\
&= \sigma^2[\boldsymbol{\Psi} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}']
\end{aligned}$$

d) Quelle est la matrice de covariance entre les résidus des *MCG* ($\boldsymbol{e}_{MCG}$) et les résidus des *MCO* ($\boldsymbol{e}_{MCO}$) ?

La covariance entre le résidus MCO $\boldsymbol{e}_{MCO} = [\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{\varepsilon}$ et le résidu MCG : $\boldsymbol{e}_{MCG} =$ $= [\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Psi}^{-1}]\boldsymbol{\varepsilon}$ s'obtient comme :

$$\begin{aligned}
E(\boldsymbol{e}_{MCO}\boldsymbol{e}'_{MCG}) &= E[(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')\boldsymbol{\varepsilon\varepsilon}'(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Psi}^{-1})']\\
&= (\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')E[\boldsymbol{\varepsilon\varepsilon}'](\boldsymbol{I} - \boldsymbol{\Psi}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X})\\
&= \sigma^2(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')\boldsymbol{\Psi}(\boldsymbol{I} - \boldsymbol{\Psi}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X})\\
&= \sigma^2\boldsymbol{\Psi} - \sigma^2\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Psi} - \sigma^2\boldsymbol{\Psi}\boldsymbol{\Psi}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}\\
&\qquad + \sigma^2\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Psi}\boldsymbol{\Psi}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}\\
&= \sigma^2\boldsymbol{\Psi} - \sigma^2\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Psi} - \sigma^2\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X} + \sigma^2\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Psi}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}\\
&= \sigma^2[\boldsymbol{I} - \sigma^2\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{\Psi}\\
&= \sigma^2\boldsymbol{M}\boldsymbol{\Psi}
\end{aligned}$$

74. **Considérons le modèle linéaire suivant pour expliquer la consommation mensuelle de bière :**

$$bière = \beta_1 + \beta_2 revenu + \beta_3 prix + \beta_4 etude + \beta_5 femme + \varepsilon$$

$$\text{avec : } \begin{cases} E(u|revenu, prix, etude, femme) = 0 \\ V(u|revenu, prix, etude, femme) = \sigma^2 revenu^2 \end{cases}$$

Ecrivez le modèle transformé avec des erreurs homoscédastiques. Interprétez ce modèle.

[Wooldridge (2009), exercice 8.2]

75. **Considérons un modèle appliqué à un ensemble d'employés :**

$$y_{i,e} = \beta_0 + \beta_1 x_{1,i,e} + \beta_2 x_{2,i,e} e + \cdots + \beta_4 x_{k,i,e} + v_{i,e} + f_i$$

où la variable non observée f_i capture un « effet d'entreprise », c'est-à-dire l'effet que les caractéristiques propres de l'entreprise *i* peuvent avoir sur la variable dépendante. Le terme d'erreur $v_{i,e}$ est spécifique à l'employé *e* de l'entreprise *i*. L'erreur composite $u_{i,e} = v_{i,e} + f_i$ est celle de l'équation (8.28).

a) **Supposons que $Var(f_i) = \sigma_f^2$,$Var(v_{i,e}) = \sigma_v^2$ et que $v_{i,e}$ et f_i ne sont pas corrélés. Montrez alors que $Var(u_{i,e}) = \sigma_f^2 + \sigma_v^2$. Appelez cette variance σ^2.**
b) **Supposons maintenant que $v_{i,e}$ et $v_{i,g}$ ne sont pas corrélés pour $e \neq g$. Montrez que $v(u_{i,e}\ ,\ u_{i,g}) = \sigma_f^2$.**
c) **Soit la moyenne des erreurs composites au sein d'une entreprise : $\overline{u_i} = m_i^{-1} \sum_{e=1}^{m_i} u_{i,e}$ avec m_i le nombre d'employés d'une entreprise *i*. Montrez que $Var(\overline{u_i}) = \sigma_f^2 + \sigma_v^2 / m_i$.**
d) **Discutez l'intérêt du point (iii) pour l'application de la méthode des MCP à des données agrégées au niveau de l'entreprise, où le poids utilisé pour l'observation *i* est fonction de la taille de l'entreprise.**

[voir Wooldridge (2009), exercice 8.7]

76. Démontrez que la statistique de Lung-Box est supérieure à la statistique de Box-Pierce. Montrez que ces deux statistiques de test sont asymptotiquement équivalentes. Qu'en concluez-vous ?

Les statistiques de Box – Pierce et de Ljung – Box permettent de tester la présence d'autocorrélation dans les erreurs du modèle.

On calcule d'abord les autocorrélations des résidus MCO du modèle $\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}_{MCO}}$:

$$\widehat{\rho_s} = \frac{\frac{1}{T}\sum_{t=s+1}^{T} e_t e_{t-s}}{\frac{1}{T}\sum_{t=1}^{T} e_t^2}$$

On a vu dans le cours que, si les régresseurs sont strictement exogènes, $\sqrt{T}\widehat{\rho_s} \overset{d}{\rightarrow} N(0,1)$, sous l'hypothèse nulle. Dès lors le test de Box – Pierce est défini par :

$$Q_L = T\sum_{s=1}^{L} \widehat{\rho_s}^2 \overset{d}{\rightarrow} \chi^2(L) \quad \text{sous } H_0$$

tandis que le test de Ljung-Box est alors :

$$Q_L' = T(T+2)\sum_{s=1}^{L} \frac{\widehat{\rho_s}^2}{T-s} \overset{d}{\rightarrow} \chi^2(L) \quad \text{sous } H_0$$

La différence entre les deux tests est alors :

$$Q_L' - Q_L = T\sum_{s=1}^{L} \frac{T+2}{T-s}\widehat{\rho_s}^2 - T\sum_{s=1}^{L} \widehat{\rho_s}^2$$

$$= T\sum_{s=1}^{L} \left(\frac{T+2}{T-s} - 1\right)\widehat{\rho_s}^2$$

Lorsque $T \rightarrow \infty$, le facteur $\left(\frac{T+2}{T-s} - 1\right) \rightarrow 0$. La différence entre les deux statistiques de test disparaît asymptotiquement. Ce qui prouve qu'ils sont asymptotiquement équivalents.

Comme $s > 0$, le facteur $\frac{T+2}{T-s}$ est supérieur à 1. Ce qui implique que $\frac{T+2}{T-s} - 1 > 0$ et que la différence entre les deux tests est positive : $Q_L' - Q_L > 0$. Donc en échantillon fini, la statistique de Ljung – Box est supérieure à la statistique de Box – Pierce : $Q_L' > Q_L$. La probabilité critique du test de Ljung – Box sera ainsi plus petite que celle du test de Box – Pierce. On rejettera plus souvent l'hypothèse nulle d'absence d'autocorrélation avec le test de Ljung – Box qu'avec celui de Box – Pierce.

77. Supposons un processus stochastique $\{y_t\}$ qui est généré par le modèle : $y_t = z + \varepsilon_t$ pour $t = 1, 2, \ldots$ avec $\{\varepsilon_t\}$ une séquence *i.i.d.* d'espérance nulle et de variance σ_ε^2. La variable aléatoire z est invariante dans le temps avec une espérance nulle et une variance σ_z^2. On suppose que chaque ε_t n'est pas corrélé avec z.

a) Trouvez l'espérance et la variance de y_t. Est-ce qu'elles dépendent du temps ?
b) Donnez la covariance : $Cov(y_t, y_{t+h})$ pour tout t et h. Est-ce que le processus $\{y_t\}$ est stationnaire en covariance ?
c) Montrez que les autocorrélations sont : $Corr(y_t, y_{t+h}) = \sigma_z^2/(\sigma_z^2 + \sigma_\varepsilon^2)$ pour tout t et h.
d) Est-ce que y_t est asymptotiquement non corrélé ? Expliquez.

[voir Wooldridge (2009), exercice 11.3]

78. Considérez le modèle : $y_t = x_t'\beta + u_t$ avec $u_t = \rho u_{t-1} + \varepsilon_t$ où ε_t est un bruit blanc gaussien. Comparez les autocorrélations de u_t dans le modèle original, avec celles de v_t dans le modèle en différence première : $(y_t - y_{t-1}) = (x_t - x_{t-1})'\beta + (u_t - u_{t-1})$ ou $\Delta y_t = \Delta x_t{}'\beta + v_t$ avec $v_t = u_t - u_{t-1}$.
Est-ce que la transformation en différence première réduit l'autocorrélation ?
[voir Greene (2018), exercice 20.1]

L'erreur est un processus autorégressif d'ordre 1 stationnaire parce que $|\rho| < 1$. En conséquence : $V(u_t) = V(u_{t-1}) = \sigma_u^2$. D'après le processus de l'erreur : $u_t = \rho u_{t-1} + \varepsilon_t$, on aura pour la variance de u_t :

$$V(u_t) = \rho^2 V(u_{t-1}) + V(\varepsilon_t) + 2\rho Cov(u_{t-1}, \varepsilon_t)$$

La covariance entre u_{t-1} et ε_t est nulle parce que u_{t-1} est une fonction des erreurs : $\varepsilon_{t-1}, \varepsilon_{t-2}, \varepsilon_{t-3}, \ldots$ et que ces erreurs sont par définition indépendantes. En conséquence :

$$\sigma_u^2 = \rho^2 \sigma_u^2 + \sigma_\varepsilon^2$$
$$(1-\rho^2)\sigma_u^2 = \sigma_\varepsilon^2$$
$$\sigma_u^2 = \frac{\sigma_\varepsilon^2}{1-\rho^2}$$

Pour la variance de $v_t = u_t - u_{t-1}$, on aura en utilisant les règles de calcul de la variance d'une combinaison linéaire de variables aléatoires :

$$V(v_t) = V(u_t) + V(u_{t-1}) - 2Cov(u_t, u_{t-1})$$

Mais la covariance entre u_t et u_{t-1} est :

$$Cov(u_t, u_{t-1}) = Cov(\rho u_{t-1} + \varepsilon_t, u_{t-1}) = \rho E(u_{t-1}^2) + E(\varepsilon_t u_{t-1})$$

Le dernier terme est nul parce que l'innovation ε_t au temps t n'est pas corrélée à u_{t-1} qui est une somme pondérée de $\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots$ En conséquence :

$$Cov(u_t, u_{t-1}) = \rho\sigma_u^2 = \frac{\rho}{1-\rho^2}\sigma_\varepsilon^2$$

On aura alors :

$$V(v_t) = V(u_t) + V(u_{t-1}) - 2Cov(u_t, u_{t-1})$$
$$V(v_t) = \sigma_u^2 + \sigma_u^2 - 2\rho\sigma_u^2 = 2(1-\rho)\sigma_u^2$$
$$\sigma_v^2 = 2(1-\rho)\sigma_u^2$$

Comme $\sigma_u^2 = \sigma_\varepsilon^2/(1-\rho^2)$, on aura :

$$\sigma_v^2 = 2(1-\rho)\frac{\sigma_\varepsilon^2}{1-\rho^2} = \frac{2(1-\rho)}{(1-\rho)(1+\rho)}\sigma_\varepsilon^2 = \frac{2}{1+\rho}\sigma_\varepsilon^2$$

On peut aussi arriver à ce résultat en notant que :

$$v_t = u_t - u_{t-1} \quad avec \;\; u_t = \rho u_{t-1} + \varepsilon_t$$

$$v_t = (\rho u_{t-1} + \varepsilon_t) - u_{t-1}$$
$$v_t = (\rho - 1)u_{t-1} + \varepsilon_t$$

On calcule alors la variance de v_t :

$$V(v_t) = (\rho-1)^2 V(u_{t-1}) + V(\varepsilon_t) + 2(\rho-1)Cov(u_{t-1}, \varepsilon_t)$$

De nouveau : $Cov(u_{t-1}, \varepsilon_t) = E(u_{t-1}\varepsilon_t) = 0$, et on obtient :

$$\sigma_v^2 = V(v_t) = (\rho-1)^2\sigma_u^2 + \sigma_\varepsilon^2$$

$$
\begin{aligned}
\sigma_v^2 &= (\rho-1)^2\left(\frac{1}{1-\rho^2}\right)\sigma_\varepsilon^2+\sigma_\varepsilon^2 \\
&= \left[\frac{(\rho-1)^2}{(1-\rho)(1+\rho)}+1\right]\sigma_\varepsilon^2 \\
&= \left[1-\frac{(\rho-1)}{(1+\rho)}\right]\sigma_\varepsilon^2 \\
&= \left[\frac{1+\rho-\rho+1}{1+\rho}\right]\sigma_\varepsilon^2
\end{aligned}
$$

Finalement :

$$\sigma_v^2 = \left(\frac{2}{1+\rho}\right)\sigma_\varepsilon^2$$

La variance de v_t est plus petite que la variance de u_t si $0{,}50 < \rho < 1$. En revanche, la variance de v_t est plus grande que la variance de u_t si $-1 < \rho < 0{,}50$. En effet :

$$\sigma_v^2 = 2(1-\rho)\sigma_u^2 < \sigma_u^2 \quad \text{si } 2(1-\rho)<1$$
$$\text{ou si } 1-\rho < \tfrac{1}{2} \text{ ou si } \rho > \tfrac{1}{2}$$

Si le processus est stationnaire et infini dans le passé, on peut réécrire l'erreur comme une suite des innovations qui déclinent géométriquement :

$$u_t = \varepsilon_t + \rho\varepsilon_{t-1} + \rho^2\varepsilon_{t-2} + \rho^3\varepsilon_{t-3} + \cdots$$

La covariance entre u_t et u_{t-s} sera alors :

$$
\begin{aligned}
Cov(u_t, u_{t-s}) = Cov(&\varepsilon_t + \rho\varepsilon_{t-1} + \rho^2\varepsilon_{t-2} + \cdots + \rho^s\varepsilon_{t-s} + \rho^{s+1}\varepsilon_{t-s-1} + \cdots, \\
&\varepsilon_{t-s} + \rho\varepsilon_{t-s-1} + \rho^2\varepsilon_{t-s-2} + \cdots)
\end{aligned}
$$

Comme les innovations ε_t sont indépendantes :

$$
\begin{aligned}
Cov(u_t, u_{t-s}) &= \rho^s E(\varepsilon_{t-s}^2) + \rho^{s+2}E(\varepsilon_{t-s-1}^2) + \rho^{s+4}E(\varepsilon_{t-s-2}^2) + \cdots \\
&= (\rho^s + \rho^{s+2} + \rho^{s+4} + \cdots)\sigma_\varepsilon^2 \\
&= \rho^s(1+\rho^2+\rho^4+\cdots)\sigma_\varepsilon^2 \\
&= \frac{\rho^s}{1-\rho^2}\sigma_\varepsilon^2 = \rho^s\sigma_u^2
\end{aligned}
$$

Les autocorrélations seront alors :

$$Corr(u_t, u_{t-s}) = \frac{Cov(u_t, u_{t-s})}{\sqrt{V(u_t) \times V(u_{t-s})}} = \frac{Cov(u_t, u_{t-s})}{V(u_t)} = \frac{\rho^s \sigma_u^2}{\sigma_u^2} = \rho^s$$

Pour l'erreur en différence première, on aura pour l'autocovariance d'ordre 1 :

$$\begin{aligned}
Cov(v_t, v_{t-1}) &= E[(u_t - u_{t-1})(u_{t-1} - u_{t-2})] \\
&= E(u_t u_{t-1}) - E(u_t u_{t-2}) - E(u_{t-1}^2) + E(u_{t-1} u_{t-2}) \\
&= \rho\sigma_u^2 - \rho^2\sigma_u^2 - \sigma_u^2 + \rho\sigma_u^2 \\
&= -\sigma_u^2(\rho^2 - 2\rho + 1) \\
&= -(\rho - 1)^2\sigma_u^2
\end{aligned}$$

On peut aussi obtenir :

$$\begin{aligned}
Cov(v_t, v_{t-1}) &= -(\rho - 1)^2\sigma_u^2 = -(\rho - 1)^2 \frac{\sigma_\varepsilon^2}{1 - \rho^2} \\
&= -\frac{(\rho - 1)^2}{(1 - \rho)(1 + \rho)}\sigma_\varepsilon^2 = \frac{\rho - 1}{1 + \rho}\sigma_\varepsilon^2
\end{aligned}$$

Ainsi l'autocorrélation d'ordre 1 sera :

$$Corr(v_t, v_{t-1}) = \frac{Cov(v_t, v_{t-1})}{\sqrt{V(v_t) \times V(v_{t-1})}} = \frac{Cov(v_t, v_{t-1})}{V(v_t)} = \frac{-(\rho - 1)^2\sigma_u^2}{2(1 - \rho)\sigma_u^2} = \frac{\rho - 1}{2}$$

On procède de même pour les autocovariances d'ordre s :

$$\begin{aligned}
Cov(v_t, v_{t-s}) &= E[(u_t - u_{t-1})(u_{t-s} - u_{t-s-1})] \\
&= E(u_t u_{t-s}) - E(u_t u_{t-s-1}) - E(u_{t-1} u_{t-s}) + E(u_{t-1} u_{t-s-1}) \\
&= \rho^s\sigma_u^2 - \rho^{s+1}\sigma_u^2 - \rho^{s-1}\sigma_u^2 + \rho^s\sigma_u^2 \\
&= -\rho^{s-1}\sigma_u^2(\rho^2 - 2\rho + 1) \\
&= -\rho^{s-1}(\rho - 1)^2\sigma_u^2
\end{aligned}$$

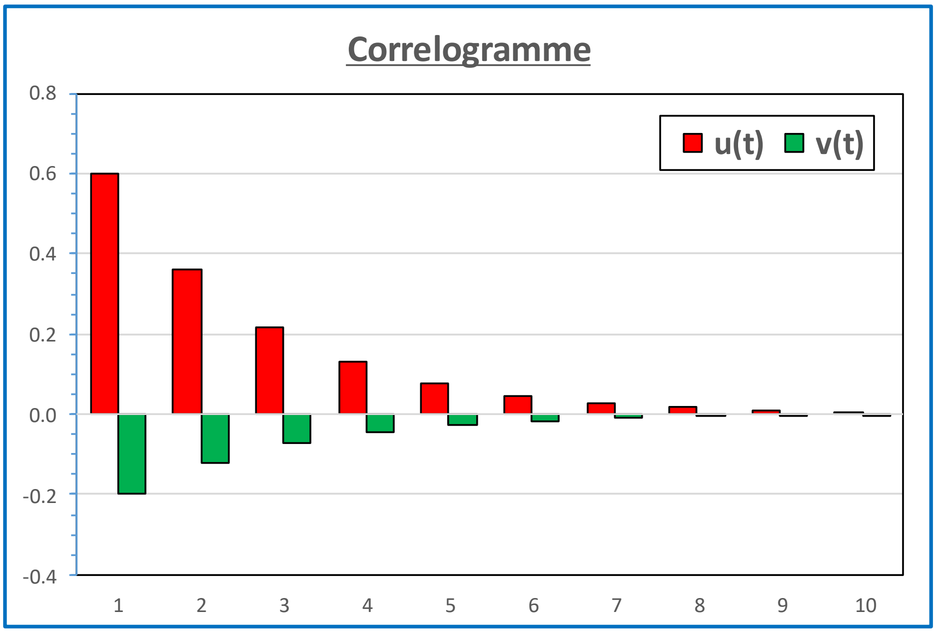
Ainsi l'autocorrélation d'ordre s sera :

$$Corr(v_t, v_{t-s}) = \frac{Cov(v_t, v_{t-s})}{\sqrt{V(v_t) \times V(v_{t-s})}} = \frac{Cov(v_t, v_{t-s})}{V(v_t)} = \frac{-\rho^{s-1}(\rho - 1)^2\sigma_u^2}{2(1 - \rho)\sigma_u^2} = \frac{\rho^{s-1}(\rho - 1)}{2}$$

Pour $\rho = 0{,}60$, on aura les autocorrélations suivantes :

s	u_t	v_t
1	0,600	-0.200
2	0,360	-0.120
3	0,216	-0.072
4	0.130	-0.043
5	0.078	-0.026
6	0.047	-0.016
Etc…		

Le corrélogramme est alors :



ÉCONOMETRIE THEORIQUE

(M1 MBFA)

Corrigés des Exercices Théoriques

Chapitre 4 : L'estimateur de la Méthode des Moments Généralisée

79. Soit un modèle d'offre – demande classique de la forme :

$$\begin{cases} \boldsymbol{q_i = \alpha_0 + \alpha_1 p_i + u_i} & \rightarrow \text{ équation de demande} \\ \boldsymbol{q_i = \beta_0 + \beta_1 p_i + v_i} & \rightarrow \text{ équation d'offre} \end{cases}$$

a) Supposons que la covariance entre les deux erreurs n'est pas nulle : $Cov(u_i, v_i) \neq 0$. est-ce que le prix et le choc de demande (u_i) sont positivement corrélés quand $\alpha_1 < 0$ et $\beta_1 > 0$?

La solution de ce système pour la quantité et le prix d'équilibre est :

$$\begin{cases} p = \dfrac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} + \dfrac{v - u}{\alpha_1 - \beta_1} \\ \\ q = \dfrac{\alpha_1 \beta_0 - \alpha_0 \beta_1}{\alpha_1 - \beta_1} + \dfrac{\alpha_1 v - \beta_1 u}{\alpha_1 - \beta_1} \end{cases}$$

Calculons la covariance entre le prix et le choc de demande :

$$\begin{aligned} Cov(p, u) &= Cov\left(\frac{v - u}{\alpha_1 - \beta_1}, u\right) \\ &= Cov\left(\frac{v}{\alpha_1 - \beta_1}, u\right) - Cov\left(\frac{u}{\alpha_1 - \beta_1}, u\right) \\ &= \frac{1}{\alpha_1 - \beta_1} Cov(v, u) - \frac{1}{\alpha_1 - \beta_1} V(u) \\ &= \frac{1}{\alpha_1 - \beta_1} [Cov(v, u) - V(u)] \end{aligned}$$

Si le deuxième terme est positif parce que $\alpha_1 - \beta_1 < 0$, le premier terme peut être positif si la covariance entre les chocs d'offre et de demande sont corrélés négativement ! Pour que la corrélation entre choc de demande et prix deviennent négatif, il faudrait que $Cov(v,u) > V(u)$.

b) On a vu que l'estimateur des MCO du paramètres du prix dans une régression de la quantité sur une constante et la prix est biaisé pour α_1. Est-ce que le paramètre estimé de la constante est aussi biaisé pour α_0 ?

Soit la régression : $q = \gamma_0 + \gamma_1 p + \varepsilon$. L'estimateur MCO de la constante est $\widehat{\gamma_0} = \bar{q} - \widehat{\gamma_1}\bar{p}$ avec $\widehat{\gamma_1} = Cov(p,q)/V(p)$. La covariance entre le prix et la quantité peut être calculée en fonction de la courbe de demande : $q = \alpha_0 + \alpha_1 p + u$:

$$Cov(p,q) = \alpha_1 V(p) + Cov(p,u)$$

Ce qui implique la non convergence de l'estimateur $\widehat{\gamma_1}$ pour α_1 :

$$\text{plim}\, \widehat{\gamma_1} = \alpha_1 + \frac{Cov(p,u)}{V(p)} \neq \alpha_1$$

Maintenant, on aura pour l'estimateur de la constante $\widehat{\gamma_0}$ avec $\bar{q} = \alpha_0 + \alpha_1\bar{p} + \bar{u}$:

$$\begin{aligned}
\text{plim}\, \widehat{\gamma_0} &= \text{plim}(\bar{q} - \widehat{\gamma_1}\bar{p}) \\
&= \text{plim}(\alpha_0 + \alpha_1\bar{p} + \bar{u}) - \text{plim}\, \widehat{\gamma_1} \times \text{plim}\, \bar{p} \\
&= \alpha_0 + \alpha_1 \text{plim}(\bar{p}) + \text{plim}(\bar{u}) - \text{plim}\, \widehat{\gamma_1} \times \text{plim}\, \bar{p}
\end{aligned}$$

Mais on aura, par une loi des grands nombres, $\text{plim}(\bar{p}) = E(p)$ et $\text{plim}(\bar{u}) = E(u) = 0$.

$$\begin{aligned}
\text{plim}\, \widehat{\gamma_0} &= \alpha_0 + \alpha_1 E(p) - \left(\alpha_1 + \frac{Cov(p,u)}{V(p)}\right) E(p) \\
&= \alpha_0 - \frac{Cov(p,u)}{V(p)} E(p) \neq \alpha_0
\end{aligned}$$

Comme l'estimateur $\widehat{\gamma_1}$ est non convergent pour α_1, le second terme n'est pas nul (sauf si $E(p) = 0$, ce qui n'est pas très envisageable !) , l'estimateur $\widehat{\gamma_0}$ sera aussi non convergent pour α_0.

c) Démontrez que l'estimateur du paramètre du prix dans une régression de la quantité sur une constante et le prix ($q_i = \gamma_0 + \gamma_1 p_i + \eta_i$) a pour limite en probabilité :

$$\textbf{plim}\ \widehat{\boldsymbol{\gamma_1}} = \frac{\boldsymbol{\alpha_1 V(v_i) + \beta_1 V(u_i)}}{\boldsymbol{V(v_i) + V(u_i)}}$$

La limite en probabilité de l'estimateur de la pente est :

$$\text{plim}\ \widehat{\gamma_1} = \frac{Cov(p,q)}{V(p)}$$

Du fait de la solution d'équilibre pour le prix est :

$$p = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} + \frac{v-u}{\alpha_1 - \beta_1}$$

la variance du prix sera :

$$V(p) = \frac{V(v) + V(u)}{(\alpha_1 - \beta_1)^2}$$

La covariance entre le prix et la quantité d'équilibre sera :

$$\begin{aligned} Cov(p,q) &= Cov\left(\frac{v-u}{\alpha_1 - \beta_1}, \frac{\alpha_1 v - \beta_1 u}{\alpha_1 - \beta_1}\right) \\ &= \frac{1}{(\alpha_1 - \beta_1)^2} Cov(v-u, \alpha_1 v - \beta_1 u) \\ &= \frac{1}{(\alpha_1 - \beta_1)^2} [\alpha_1 V(v) - \alpha_1 Cov(u,v) - \beta_1 Cov(u,v) - \beta_1 V(u)] \end{aligned}$$

Comme on a supposé que les chocs d'offre et de demande n'étaient pas corrélés $Cov(u,v) = 0$:

$$Cov(p,q) = \frac{\alpha_1 V(v) - \beta_1 V(u)}{(\alpha_1 - \beta_1)^2}$$

Ce qui donne pour la limité en probabilité de l'estimateur de la pente :

$$\text{plim}\ \widehat{\gamma_1} = \frac{Cov(p,q)}{V(p)} = \frac{\dfrac{\alpha_1 V(v) - \beta_1 V(u)}{(\alpha_1 - \beta_1)^2}}{\dfrac{V(v) + V(u)}{(\alpha_1 - \beta_1)^2}} = \frac{\alpha_1 V(v) - \beta_1 V(u)}{V(v) + V(u)}$$

80. Sous quelles hypothèses l'estimateur par variables instrumentales est convergent ? Démontrez la convergence de cet estimateur. Donnez la distribution asymptotique de l'estimateur VI.

81. Montrez que la régression par MCO est un cas particulier de l'estimation par variables instrumentales.

82. Soit $\boldsymbol{\Sigma}_{xz} = E(\boldsymbol{x}_i \boldsymbol{z}_i')$ de dimension $K \times L$ et $\boldsymbol{\Sigma}_{xy} = E(\boldsymbol{x}_i y_i)$ de dimension $K \times 1$. Montrez que :

$$\textbf{rang}(\boldsymbol{\Sigma}_{xz}) = \textbf{rang}(\boldsymbol{\Sigma}_{xz} \ \vdots \ \boldsymbol{\Sigma}_{xy})$$

L'estimateur des GMM repose sur les conditions d'orthogonalité :

$$E(\boldsymbol{g}_i) = E(\boldsymbol{x}_i \varepsilon_i) = E\left(\boldsymbol{x}_i\left(y_i - \boldsymbol{z}_i'\widetilde{\boldsymbol{\delta}}\right)\right) = \mathbf{0}$$

que l'on peut décomposer :

$$E(\boldsymbol{x}_i y_i) - E(\boldsymbol{x}_i \boldsymbol{z}_i')\widetilde{\boldsymbol{\delta}} = \mathbf{0} \quad \Rightarrow \quad \boldsymbol{\Sigma}_{xy} = \boldsymbol{\Sigma}_{xz}\widetilde{\boldsymbol{\delta}}$$

$\boldsymbol{\Sigma}_{xy}$ est une combinaison linéaire des L colonnes de $\boldsymbol{\Sigma}_{xz}$. En conséquence, si on ajoute $\boldsymbol{\Sigma}_{xy}$ aux colonnes de $\boldsymbol{\Sigma}_{xz}$, le rang reste le même :

$$\text{rang}(\boldsymbol{\Sigma}_{xz}) = \text{rang}(\boldsymbol{\Sigma}_{xz} \ \vdots \ \boldsymbol{\Sigma}_{xy}) = L$$

83. Si on ajoute une autre variable instrumentale (ξ_i) aux instruments $(\boldsymbol{x}_i)$. Bien que cette variable supplémentaire soit prédéterminée, elle n'est pas reliée aux régresseurs : $E(\xi_i z_{i,l}) = 0$ pour tout $l = 1, 2, \dots, L$. Est-ce que la condition de rang est toujours satisfaite ?
[Hayashi (2000), Exercices 3.3(6)]

La condition de rang pour l'identification requiert que la matrice $\boldsymbol{\Sigma}_{xz}$ soit de rang plein colonne $\text{rang}(\boldsymbol{\Sigma}_{xz}) = L \leq K$. Si on ajoute une ligne supplémentaire à cette matrice avec une autre variable instrumentale, cela ne change pas le rang de cette matrice. En conséquence, la condition de rang est toujours satisfaite.

84. Soit la matrice A de dimension $q \times K$ de rang plein ligne avec $q \leq K$ telle que $A\Sigma_{xz}$ est de rang plein colonne. Notons $\hat{x}_i = Ax_i$. Vérifiez que les hypothèses 3.3, 3.4 et 3.5 tiennent pour les instruments transformés $\hat{x}_i$ si elles sont satisfaites pour les variables instrumentales x_i.
[Hayashi (2000), Exercices 3.3(8)]

85. Si l'équation est juste identifiée, quelle est la valeur minimisée de $J(\tilde{\delta}, \widehat{W})$?
[Hayashi (2000), Exercices 3.4(2)]

86. Soit un modèle de régression simple $y_t = \alpha + \beta z_t + \varepsilon_t$ avec $\varepsilon_t \approx i.i.d.(0, \sigma_\varepsilon^2)$. On suppose que $E(z_t \varepsilon_t) \neq 0$, mais qu'on a une variable instrumentale x_t corrélée avec z_t telle que $E(x_t \varepsilon_t) = 0$.

Montrez que l'estimateur par variable instrumentale de β est équivalent à l'estimateur par MCO de β dans modèle : $y_t = \alpha + \beta z_t + \gamma v_t + \varepsilon_t$ où v_t est le résidus de la régression de première étape : $z_t = \delta + \theta x_t + v_t$.
[Proposition de Durbin (1954) et Wu (1973)]

Réécrivons matriciellement la régression de première étape comme : $\boldsymbol{Z} = \boldsymbol{X\theta} + \boldsymbol{v}$, dont le résidus de l'estimation des MCO est $\boldsymbol{v} = \boldsymbol{Z} - \widehat{\boldsymbol{Z}} = \boldsymbol{M_x Z}$ avec $\boldsymbol{M_x} = \boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$.

Avec le théorème de Frisch-Waugh, le paramètre β de la régression augmentée du résidu de première étape $\boldsymbol{y} = \boldsymbol{Z\beta} + \boldsymbol{v\gamma} + \boldsymbol{\varepsilon}$ peut être estimé par MCO sous la forme :

$$\widetilde{\boldsymbol{\beta}_{MCO}} = (\boldsymbol{Z}'\boldsymbol{M_v}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{M_v}\boldsymbol{y}$$

avec $\boldsymbol{M_v} = \boldsymbol{I} - \boldsymbol{v}(\boldsymbol{v}'\boldsymbol{v})^{-1}\boldsymbol{v}'$. Mais $\boldsymbol{v} = \boldsymbol{M_x Z}$, ce qui donne :

$$\boldsymbol{M_v} = \boldsymbol{I} - \boldsymbol{M_x Z}(\boldsymbol{Z}'\boldsymbol{M_x Z})^{-1}\boldsymbol{Z}'\boldsymbol{M_x}$$

En substituant dans l'expression de $\widetilde{\boldsymbol{\beta}_{MCO}}$, on obtient :

$$
\begin{aligned}
\widetilde{\beta_{MCO}} &= [Z'(I - M_x z(z'M_x z)^{-1} z'M_x)Z]^{-1} Z'(I - M_x z(z'M_x z)^{-1} z'M_x)y \\
&= [Z'Z - Z'M_x Z(Z'M_x Z)^{-1} Z'M_x Z]^{-1}(Z'y - Z'M_x Z(Z'M_x Z)^{-1} Z'M_x y) \\
&= [Z'Z - Z'M_x Z]^{-1}(Z'y - Z'M_x y) \\
&= (Z'P_x Z)^{-1} Z'P_x y
\end{aligned}
$$

avec $P_x = I - M_x = X(X'X)^{-1}X'$, parce que $Z'Z - Z'M_x Z = Z'(I - M_x)Z = Z'P_x Z$ et de la même manière : $Z'y - Z'M_x y = Z'P_x y$. Maintenant remplaçons P_x dans l'expression précédente :

$$
\begin{aligned}
\widetilde{\beta_{MCO}} &= (Z'P_x Z)^{-1} Z'P_x y \\
&= (Z'X(X'X)^{-1}X'Z)^{-1} Z'X(X'X)^{-1}X'y \\
&= \widetilde{\beta_{DMC}}
\end{aligned}
$$

qui est l'estimateur des DMC du modèle initial $y = Z\beta + \varepsilon$ avec les instruments X.

87. Supposons que l'on recueille par une enquête des données sur la consommation de vin des ménages, ainsi que leurs revenus. On veut estimer l'élasticité de la consommation de vin au revenu. Mais ces deux variables sont mal renseignées, ou avec des erreurs.

a) Quelle est la conséquence des erreurs de mesure sur la variable dépendante (consommation de vin) et sur la variable explicative (le revenu) ?

b) Un économètre propose d'utiliser la variable instrumentale : montant total des chèques émis par le ménage. Pensez-vous que cette variable est un bon instrument ? Pourquoi ?

88. Dans un modèle de détermination des salaires, on explique souvent le log ($salaire$) par l'éducation mesurée par le nombre d'années d'étude ($etude$) et par d'autres variables explicatives X. Soit le modèle :

$$\log(salaire) = \beta_1 + \beta_2 etude + X\gamma + \varepsilon$$

Or le nombre d'année d'étude ne mesure qu'imparfaitement la qualité de la formation suivie par la personne. Quel est alors la conséquence de prendre la variable $etude$ dans cette régression au lieu de la qualité de la formation ?

89. Dans un modèle où la variable explicative est endogène, l'estimateur des MCO est biaisé et non convergent : plim $\widehat{\beta_{MCO}} = \beta + Q^{-1}\gamma = \theta$ avec $Q = \text{plim}(X'X/N)$ et $\gamma = \text{plim}(X'\varepsilon/N)$.

a) Dérivez la matrice de variance-covariance asymptotique de l'estimateur $\widehat{\beta_{MCO}}$.

b) Montrez que l'estimateur $\widehat{\beta_{MCO}}$ est asymptotiquement normalement distribué.

[Greene (2018) : Exercice 8.1]

90. Supposons deux choix de matrice de pondération $\widehat{W_1}$ et $\widehat{W_2}$ tel que $\widehat{W_1} - \widehat{W_2} \overset{p}{\to} 0$. Démontrez que :

$$\sqrt{N}\,\hat{\delta}(\widehat{W_1}) - \sqrt{N}\,\hat{\delta}(\widehat{W_2}) \overset{p}{\to} 0$$

[Hayashi (2000), Exercices 3.5(3)]

L'estimateur des GMM s'écrit :

$$\hat{\delta}(\widehat{W_l}) = \left(S'_{xz}\widehat{W_l}S_{xz}\right)^{-1} S'_{xz}\widehat{W_l}S_{xy}$$

Mais à partir du modèle $y_i = z_i'\delta + \varepsilon_i$, on a : $S_{xy} = S_{xz}\delta + \bar{g}$ avec $\bar{g} = \frac{1}{N}\sum_{i=1}^{N} x_i\varepsilon_i$. En conséquence :

$$\hat{\delta}(\widehat{W_l}) = \left(S'_{xz}\widehat{W_l}S_{xz}\right)^{-1} S'_{xz}\widehat{W_l}(S_{xz}\delta + \bar{g}) = \delta + \left(S'_{xz}\widehat{W_l}S_{xz}\right)^{-1} S'_{xz}\widehat{W_l}\bar{g}$$

Donc la différence des deux estimateurs sera :

$$\begin{aligned}
\sqrt{N}\,\hat{\delta}(\widehat{W_1}) - \sqrt{N}\,\hat{\delta}(\widehat{W_2}) &= \sqrt{N}\left[\hat{\delta}(\widehat{W_1}) - \hat{\delta}(\widehat{W_1})\right] \\
&= \sqrt{N}\left[\left(S'_{xz}\widehat{W_1}S_{xz}\right)^{-1} S'_{xz}\widehat{W_1}\bar{g} - \left(S'_{xz}\widehat{W_2}S_{xz}\right)^{-1} S'_{xz}\widehat{W_2}\bar{g}\right] \\
&= \left[\left(S'_{xz}\widehat{W_1}S_{xz}\right)^{-1} S'_{xz}\widehat{W_1} - \left(S'_{xz}\widehat{W_2}S_{xz}\right)^{-1} S'_{xz}\widehat{W_2}\right]\sqrt{N}\bar{g}
\end{aligned}$$

Avec les hypothèses asymptotiques habituelles $\sqrt{N}\bar{g}$ converge vers une variable aléatoire normalement distribuée. Mais

$$\text{plim}\left[\left(\boldsymbol{S}'_{xz}\widehat{\boldsymbol{W}}_{\mathbf{1}}\boldsymbol{S}_{xz}\right)^{-1}\boldsymbol{S}'_{xz}\widehat{\boldsymbol{W}}_{\mathbf{1}} - \left(\boldsymbol{S}'_{xz}\widehat{\boldsymbol{W}}_{\mathbf{2}}\boldsymbol{S}_{xz}\right)^{-1}\boldsymbol{S}'_{xz}\widehat{\boldsymbol{W}}_{\mathbf{2}}\right]$$
$$= \text{plim}\left(\boldsymbol{S}'_{xz}\widehat{\boldsymbol{W}}_{\mathbf{1}}\boldsymbol{S}_{xz}\right)^{-1}\boldsymbol{S}'_{xz}\widehat{\boldsymbol{W}}_{\mathbf{1}} - \text{plim}\left(\boldsymbol{S}'_{xz}\widehat{\boldsymbol{W}}_{\mathbf{2}}\boldsymbol{S}_{xz}\right)^{-1}\boldsymbol{S}'_{xz}\widehat{\boldsymbol{W}}_{\mathbf{2}}$$
$$= (\boldsymbol{\Sigma}'_{xz}\boldsymbol{W}\boldsymbol{\Sigma}_{xz})^{-1}\boldsymbol{\Sigma}'_{xz}\boldsymbol{W} - (\boldsymbol{\Sigma}'_{xz}\boldsymbol{W}\boldsymbol{\Sigma}_{xz})^{-1}\boldsymbol{\Sigma}'_{xz}\boldsymbol{W} = \mathbf{0}$$

si $\widehat{\boldsymbol{W}}_{\mathbf{1}} \xrightarrow{p} \boldsymbol{W}$ et $\widehat{\boldsymbol{W}}_{\mathbf{2}} \xrightarrow{p} \boldsymbol{W}$ parce que $\widehat{\boldsymbol{W}}_{\mathbf{1}} - \widehat{\boldsymbol{W}}_{\mathbf{2}} \xrightarrow{p} \boldsymbol{W} - \boldsymbol{W} = \mathbf{0}$.

En utilisant le Lemme 2.4(b) : si $x_n \xrightarrow{d} x$ et si $y_n \xrightarrow{p} 0$, alors $y_n x_n \xrightarrow{p} 0$, on aura :

$$\sqrt{N}\,\widehat{\boldsymbol{\delta}}\left(\widehat{\boldsymbol{W}}_{\mathbf{1}}\right) - \sqrt{N}\,\widehat{\boldsymbol{\delta}}\left(\widehat{\boldsymbol{W}}_{\mathbf{2}}\right) \xrightarrow{p} \mathbf{0}$$

91. Démontrez que :

$$\boldsymbol{J}\left(\widehat{\boldsymbol{\delta}}\left(\widehat{\boldsymbol{S}}^{-1}\right), \widehat{\boldsymbol{S}}^{-1}\right) = \boldsymbol{N}\boldsymbol{s}'_{xy}\widehat{\boldsymbol{S}}^{-1}\left(\boldsymbol{s}_{xy} - \boldsymbol{S}_{xz}\widehat{\boldsymbol{\delta}}\left(\widehat{\boldsymbol{S}}^{-1}\right)\right)$$

[Hayashi (2000), Exercices 3.6(2)]

Reprenons la définition de la distance à minimiser pour l'estimateur GMM :

$$J\left(\widetilde{\boldsymbol{\delta}}, \widehat{\boldsymbol{W}}\right) = N g_N\left(\widetilde{\boldsymbol{\delta}}\right)'\widehat{\boldsymbol{W}} g_N\left(\widetilde{\boldsymbol{\delta}}\right)$$

avec $g_N\left(\widetilde{\boldsymbol{\delta}}\right) = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_i\varepsilon_i = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_i\left(y_i - \boldsymbol{z}'_i\widetilde{\boldsymbol{\delta}}\right)$. On peut réécrire cette écart :

$$g_N\left(\widetilde{\boldsymbol{\delta}}\right) = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_i y_i - \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_i\boldsymbol{z}'_i\widetilde{\boldsymbol{\delta}} = \boldsymbol{s}_{xy} - \boldsymbol{S}_{xz}\widetilde{\boldsymbol{\delta}}$$

On aura ici avec le choix optimal pour la matrice de pondération $\widehat{\boldsymbol{W}} = \widehat{\boldsymbol{S}}^{-1}$:

$$J\left(\widehat{\boldsymbol{\delta}}\left(\widehat{\boldsymbol{S}}^{-1}\right), \widehat{\boldsymbol{S}}^{-1}\right) = N g_N\left(\widehat{\boldsymbol{\delta}}\left(\widehat{\boldsymbol{S}}^{-1}\right)\right)'\widehat{\boldsymbol{S}}^{-1} g_N\left(\widehat{\boldsymbol{\delta}}\left(\widehat{\boldsymbol{S}}^{-1}\right)\right)$$
$$= N\left(\boldsymbol{s}_{xy} - \boldsymbol{S}_{xz}\widehat{\boldsymbol{\delta}}\left(\widehat{\boldsymbol{S}}^{-1}\right)\right)'\widehat{\boldsymbol{S}}^{-1}\left(\boldsymbol{s}_{xy} - \boldsymbol{S}_{xz}\widehat{\boldsymbol{\delta}}\left(\widehat{\boldsymbol{S}}^{-1}\right)\right)$$
$$= N\boldsymbol{s}'_{xy}\widehat{\boldsymbol{S}}^{-1}\left(\boldsymbol{s}_{xy} - \boldsymbol{S}_{xz}\widehat{\boldsymbol{\delta}}\left(\widehat{\boldsymbol{S}}^{-1}\right)\right) - \widehat{\boldsymbol{\delta}}\left(\widehat{\boldsymbol{S}}^{-1}\right)'\boldsymbol{S}'_{xz}\widehat{\boldsymbol{S}}^{-1}\boldsymbol{s}_{xy} + \widehat{\boldsymbol{\delta}}\left(\widehat{\boldsymbol{S}}^{-1}\right)'\boldsymbol{S}'_{xz}\widehat{\boldsymbol{S}}^{-1}\boldsymbol{S}_{xz}\widehat{\boldsymbol{\delta}}\left(\widehat{\boldsymbol{S}}^{-1}\right)$$

Mais l'estimateur GMM est défini par :

$$\hat{\delta}(\hat{S}^{-1}) = \left(S'_{xz}\hat{S}^{-1}S_{xz}\right)^{-1}S'_{xz}\hat{S}^{-1}s_{xy}$$

Ce qui donne :

$$\begin{aligned}J\left(\hat{\delta}(\hat{S}^{-1}),\hat{S}^{-1}\right) &= Ns'_{xy}\hat{S}^{-1}\left(s_{xy} - S_{xz}\hat{\delta}(\hat{S}^{-1})\right) - \hat{\delta}(\hat{S}^{-1})'S'_{xz}\hat{S}^{-1}s_{xy} \\ &\quad + \hat{\delta}(\hat{S}^{-1})'S'_{xz}\hat{S}^{-1}S_{xz}\left(S'_{xz}\hat{S}^{-1}S_{xz}\right)^{-1}S'_{xz}\hat{S}^{-1}s_{xy} \\ &= Ns'_{xy}\hat{S}^{-1}\left(s_{xy} - S_{xz}\hat{\delta}\left(\hat{S}^{-1}\right)\right)\end{aligned}$$

parce que les deux dernier termes s'annulent.

92. En cas d'homoscédasticité conditionnelle, la statistique de test de Sargan est :

$$Sargan = N\frac{\left(s_{xy} - S_{xz}\widehat{\delta_{DMC}}\right)'S_{xx}^{-1}\left(s_{xy} - S_{xz}\widehat{\delta_{DMC}}\right)}{\widehat{\sigma^2}}$$

Démontrez que cette statistique est égale à $N \times R^2_{nc}$, le coefficient de détermination non centré d'une régression de $e_{DMC} = y - Z\widehat{\delta_{DMC}}$ sur X.
[Hayashi (2000), Exercices 3.8(6)]

93. Considérez le modèle de régression : $y_i = \alpha + \beta x_i^* + \varepsilon_i$. Mais la variable x_i^* n'est pas observée, on observe seulement une variable avec erreurs : $x_i = x_i^* + \nu_i$. Cependant on dispose d'une indicatrice permettant de classer les observations : $d_i = 1$ ou 0 .

a) Montrez que l'estimateur par variables instrumentales utilisant cette indicatrice comme variable instrumentale est :

$$\widehat{\beta_{VI}} = \frac{\overline{y_{d=1}} - \overline{y_{d=0}}}{\overline{x_{d=1}} - \overline{x_{d=0}}}$$

où $\overline{y_{d=1}}$ est la moyenne de y pour les observations avec $d_i = 1$, et ainsi de suite. Interprétez cette expression.

b) Quelle est la condition pour laquelle cet estimateur est calculable, interprétez cette condition ?

c) **Quel est l'expression pour l'estimateur $\widehat{\alpha_{VI}}$?**

[voir Greene (2018) : Exemple 8.2 et Exercice 8.5, ou Wooldridge(2009) : Exercice 15.3]

94. Considérez le modèle de régression simple sur séries temporelles où la variable explicative est sujette à des erreurs de mesure :

$$y_t = \alpha + \beta x_t^* + \varepsilon_t$$
$$x_t = x_t^* + \nu_t$$

a) **Introduisez $x_t^* = x_t - \nu_t$ dans le modèle et montrez que le terme d'erreur de la nouvelle équation (u_t) est négativement corrélé avec x_t si $\beta > 0$. Quelle est l'implication pour l'estimateur des MCO de β dans la régression de y sur x ?**

b) **En plus des précédentes hypothèses, on suppose que ε_t et ν_t ne sont pas corrélés avec toutes les valeurs passées de x_t^* et de ν_t, et en particulier avec x_{t-1}^* et ν_{t-1}. Montrez alors que $E(x_{t-1}u_t) = 0$.**

c) **Est-ce que x_t et x_{t-1} sont probablement corrélés ? Expliquez.**

d) **Donnez une stratégie pour estimer de manière convergente α et β dans ce modèle.**

[Wooldridge(2009) : Exercice 15.11]

<u>ÉCONOMETRIE THEORIQUE</u>

(M1 MBFA)

<u>Corrigés des Exercices Théoriques</u>

<u>*Chapitre 5 : La méthode du maximum de vraisemblance (MV)*</u>

95. Considérons un échantillon aléatoire de 10 observations :

$$\mathbf{5\ ;\ 0\ ;\ 1\ ;\ 1\ ;\ 0\ ;\ 3\ ;\ 2\ ;3\ ;4\ ;1}$$

provenant d'une distribution de Poisson. La fonction de densité de chaque observation est :

$$f(y_i|\theta) = \frac{e^{-\theta}\theta^{y_i}}{y_i!} \quad \text{avec } y_i = 0, 1, 2, \ldots \ \text{ et } \ \theta > 0$$

Pour une distribution de Poisson : $E(y_i) = V(y_i) = \theta > 0$.

a) Quelle est la vraisemblance de cet échantillon en fonction du paramètre θ.

La fonction de vraisemblance s'écrit comme le produit des fonctions de densité marginale parce que les observations sont indépendantes.

$$L(\theta|\mathbf{y}) = \prod_{i=1}^{N} f(y_i|\theta) = \prod_{i=1}^{N} \frac{e^{-\theta}\theta^{y_i}}{y_i!} = \frac{e^{-N\theta}\theta^{\sum_{i=1}^{N} y_i}}{\prod_{i=1}^{N} y_i!}$$

Dans notre exemple : $N = 10$, $\sum_{i=1}^{N} y_i = 20$ et :

$$\prod_{i=1}^{N} y_i! = 5! \times 0! \times 1! \ \times 1! \ \times 0! \ \times 3! \times 2! \times 3! \times 4! \times 1!$$

$$= 120 \times 1 \times 1 \ \times 1 \ \times 1 \ \times 6 \times 2 \times 6 \times 24 \times 1 = 207\ 360$$

Ce qui donne :

$$L(\theta|\boldsymbol{y}) = \frac{e^{-10\theta}\theta^{20}}{207\ 360}$$

Cela donne pour chaque valeur de θ, la probabilité d'avoir observé l'échantillon particulier , en supposant que les valeurs de y_i proviennent d'un tirage aléatoire dans une loi de Poisson.

b) Donnez en général pour des observations suivant une loi de Poisson, la log-vraisemblance et l'équation de vraisemblance pour l'estimateur du maximum de vraisemblance.

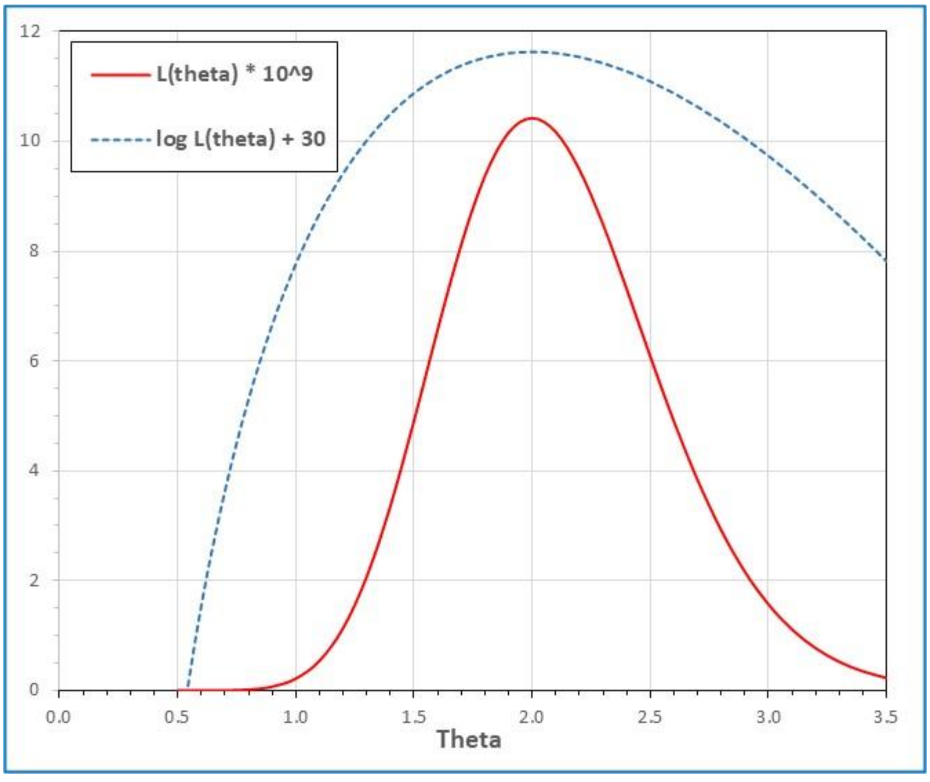
On prend le logarithme de la fonction de vraisemblance :

$$\ell(\theta|\boldsymbol{y}) = \log L(\theta|\boldsymbol{y}) = \log\left(\frac{e^{-N\theta}\theta^{\sum_{i=1}^{N} y_i}}{\prod_{i=1}^{N} y_i!}\right)$$

$$= \log\left(e^{-N\theta}\right) + \log\left(\theta^{\sum_{i=1}^{N} y_i}\right) - \log\left(\prod_{i=1}^{N} y_i!\right)$$

$$= -N\theta + \log(\theta)\sum_{i=1}^{N} y_i - \sum_{i=1}^{N} \log y_i!$$

Pour notre exemple, on donne le graphique de la fonction de vraisemblance et de la fonction de log-vraisemblance



Les équations de vraisemblance sont les dérivées premières de la log-vraisemblance par rapport aux paramètres à estimer. Il y a une seule équation de vraisemblance ici parce qu'il y a un seul paramètre à estimer.

$$\frac{\partial \ell(\theta|\boldsymbol{y})}{\partial \theta} = -N + \frac{1}{\tilde{\theta}} \sum\nolimits_{i=1}^{N} y_i = 0$$

Remarquez que cela ne dépend pas de la somme des log-factorielles.

c) Calculez l'estimateur du maximum de vraisemblance de $\boldsymbol{\theta}$. Vérifiez que vous avez bien un maximum de la log-vraisemblance.

L'estimateur du maximum de vraisemblance $\tilde{\theta}$ est alors la moyenne des valeurs observées de y_i :

$$\tilde{\theta} = \frac{1}{N} \sum\nolimits_{i=1}^{N} y_i = \bar{y}$$

Dans notre exemple,

$$\tilde{\theta} = \frac{20}{10} = 2.$$

On peut vérifier qu'on obtient bien un maximum de la fonction de log-vraisemblance en calculant la dérivée seconde :

$$\frac{\partial^2 \ell(\theta|\boldsymbol{y})}{\partial \theta^2} = \frac{-1}{\tilde{\theta}^2} \sum\nolimits_{i=1}^{N} y_i < 0$$

parce que les valeurs de y_i sont non négatives ($y_i \geq 0$) dans une loi de Poisson. Ici on aura :

$$\frac{\partial^2 \ell(\theta|\boldsymbol{y})}{\partial \theta^2} = \frac{-1}{2^2} \times 20 = -5 < 0$$

La dérivée seconde étant négative, on obtient bien un maximum qui est global.

d) Donnez la probabilité estimée des valeurs de 0 à 5.

On reprend la fonction de densité :

$$\Pr\left(y|\tilde{\theta}\right) = f(y_i|\theta) = \frac{e^{-\theta}\theta^{y_i}}{y_i!}$$

et on calcule pour $\tilde{\theta} = 2$ et $y = 0,1,2,3,4$ et 5 (avec $\exp(-2) = 0.135\,335$) :

$$\Pr\big(y = 0|\tilde{\theta} = 2\big) = \frac{e^{-2}2^0}{0!} = 0.135\,335 \times \frac{1}{1} = 0.135\,335 = 13.53\%$$

$$\Pr\big(y = 1|\tilde{\theta} = 2\big) = \frac{e^{-2}2^1}{1!} = 0.135\,335 \times \frac{2}{1} = 0.270\,671 = 27.07\%$$

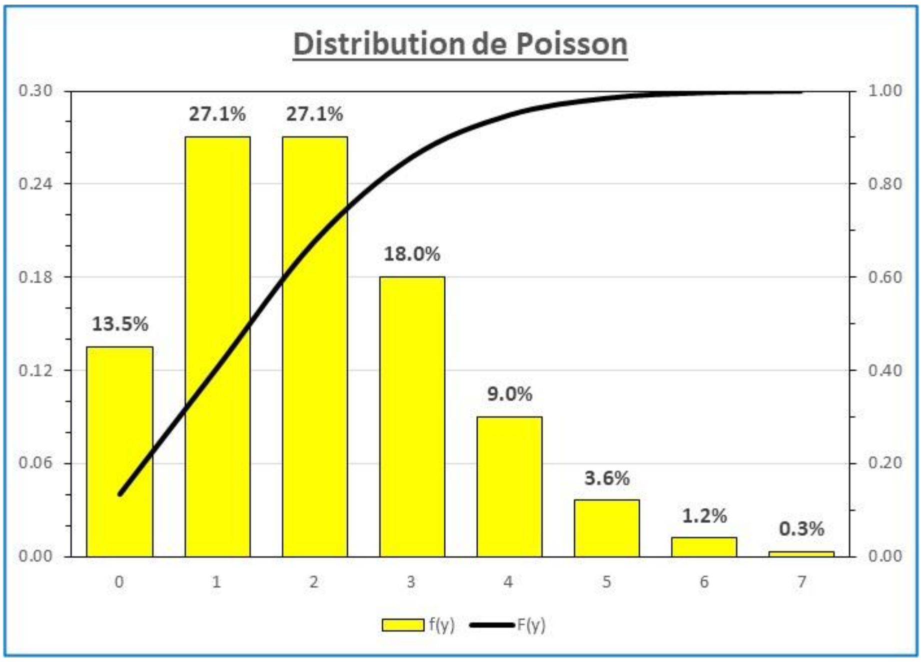
$$\Pr\big(y = 2|\tilde{\theta} = 2\big) = \frac{e^{-2}2^2}{2!} = 0.135\,335 \times \frac{4}{2} = 0.270\,671 = 27.07\%$$

$$\Pr\big(y = 3|\tilde{\theta} = 2\big) = \frac{e^{-2}2^3}{3!} = 0.135\,335 \times \frac{8}{6} = 0.180\,447 = 18.04\%$$

$$\Pr\big(y = 4|\tilde{\theta} = 2\big) = \frac{e^{-2}2^4}{4!} = 0.135\,335 \times \frac{16}{24} = 0.090\,224 = 9.02\%$$

$$\Pr\big(y = 5|\tilde{\theta} = 2\big) = \frac{e^{-2}2^5}{5!} = 0.135\,335 \times \frac{32}{120} = 0.036\,089 = 3.61\%$$

Remarquez que la somme de ces probabilités fait déjà 98.34% de la distribution. On peut aussi calculer la médiane qui est ici égale à 2. On aura alors le graphique des fonctions de densité et de distribution :



96. Dans un modèle Probit, quel est l'effet marginal d'une variable explicative sur la probabilité de choix ?

a) Si la variable explicative est continue

b) Si la variable explicative est une indicatrice.

97. On a un échantillon de disques durs d'ordinateurs dont on a mesuré la durée de vie $\boldsymbol{y_i}$ avant de tomber en panne. On va supposer que cette variable dépendante $\boldsymbol{y_i}$ est distribuée selon une loi exponentielle. Celle-ci permet de mesurer la durée de vie d'un phénomène dont le taux de panne à un moment est constant dans le temps. Sa fonction de densité est :

$$\boldsymbol{f(y_i) = \frac{1}{\mu_i} \exp\left(\frac{-y_i}{\mu_i}\right) \quad \text{pour } y_i > 0}$$

où $\boldsymbol{\mu_i = E(y_i) > 0}$ est l'espérance de $\boldsymbol{y_i}$. Par définition, sa variance est $\boldsymbol{V(y_i) = \mu_i^2}$. On suppose que $\boldsymbol{\mu_i = \exp(x_i'\beta) > 0}$ pour assurer la positivité.

a) Donnez la fonction de log-vraisemblance pour un échantillon aléatoire de $\boldsymbol{N}$ observations indépendantes.

La fonction de vraisemblance s'écrit comme le produit des fonctions de densité marginale parce que les observations sont indépendantes.

$$L(\boldsymbol{\beta}|\boldsymbol{y},\boldsymbol{x}) = \prod_{i=1}^{N} f(y_i|\theta) = \prod_{i=1}^{N} \frac{1}{\mu_i} \exp\left(\frac{-y_i}{\mu_i}\right)$$

La fonction de log-vraisemblance est alors :

$$\begin{aligned}
\ell(\boldsymbol{\beta}|\boldsymbol{y},\boldsymbol{x}) = \log L(\boldsymbol{\beta}|\boldsymbol{y},\boldsymbol{x}) &= \log\left(\prod_{i=1}^{N} \frac{1}{\mu_i} \exp\left(\frac{-y_i}{\mu_i}\right)\right) \\
&= \sum_{i=1}^{N} \log\left[\frac{1}{\mu_i} \exp\left(\frac{-y_i}{\mu_i}\right)\right] \\
&= \sum_{i=1}^{N} \left\{\log(-\mu_i) - \frac{y_i}{\mu_i}\right\}
\end{aligned}$$

Sachant que $\mu_i = \exp(\boldsymbol{x}_i'\boldsymbol{\beta})$, on aura :

$$\ell(\boldsymbol{\beta}|\boldsymbol{y},\boldsymbol{x}) = \sum_{i=1}^{N} \left\{ \log(-\exp(\boldsymbol{x}_i'\boldsymbol{\beta})) - \frac{y_i}{\exp(\boldsymbol{x}_i'\boldsymbol{\beta})} \right\}$$

$$= \sum_{i=1}^{N} \{-\boldsymbol{x}_i'\boldsymbol{\beta} - y_i \exp(-\boldsymbol{x}_i'\boldsymbol{\beta})\}$$

b) Donnez les équations de vraisemblance.

Il faut dériver la fonction de log-vraisemblance par rapport au vecteur de paramètres $\boldsymbol{\beta}$

$$\frac{\partial \ell(\boldsymbol{\beta}|\boldsymbol{y},\boldsymbol{x})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{N} \{-\boldsymbol{x}_i + (y_i \exp(-\boldsymbol{x}_i'\boldsymbol{\beta}))\boldsymbol{x}_i\}$$

$$= -\sum_{i=1}^{N} (1 - y_i \exp(-\boldsymbol{x}_i'\boldsymbol{\beta}))\boldsymbol{x}_i = \boldsymbol{0}$$

c) Peut-on donner une expression analytique pour l'estimateur du maximum de vraisemblance de $\boldsymbol{\beta}$?

Non il n'y a pas de solution analytique à ces équations de vraisemblance, parce qu'elles sont non linéaires.

d) Donnez une expression pour la matrice de variance-covariance asymptotique de l'estimateur du maximum de vraisemblance.

Il faut maintenant calculer la matrice Hessienne de la fonction de log-vraisemblance :

$$\frac{\partial^2 \ell(\boldsymbol{\beta}|\boldsymbol{y},\boldsymbol{x})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = \frac{\partial}{\partial \boldsymbol{\beta}'} \sum_{i=1}^{N} \{-\boldsymbol{x}_i + (y_i \exp(-\boldsymbol{x}_i'\boldsymbol{\beta}))\boldsymbol{x}_i\}$$

$$= -\sum_{i=1}^{N} (y_i \exp(-\boldsymbol{x}_i'\boldsymbol{\beta}))\boldsymbol{x}_i\boldsymbol{x}_i'$$

Pour des observations provenant d'un tirage aléatoire dans une loi exponentielle $y_i \geq 0$. En conséquence le facteur $y_i \exp(-\boldsymbol{x}_i'\boldsymbol{\beta}) \geq 0$, ce qui implique que la matrice Hessienne est définie-négative, assurant un estimateur du maximum de vraisemblance unique parce que la fonction de log-vraisemblance est globalement concave.

La matrice de variance – covariance asymptotique de l'estimateur du maximum de vraisemblance peut alors être estimé par la première méthode :

$$\widehat{Avar}(\tilde{\boldsymbol{\beta}}) = \left[\sum_{i=1}^{N}\left(y_i \exp\left(-\boldsymbol{x}_i'\tilde{\boldsymbol{\beta}}\right)\right)\boldsymbol{x}_i\boldsymbol{x}_i'\right]^{-1}$$

98. Supposons que la variable x a une distribution de Weibull telle que la fonction de densité est : $f(x) = \alpha\beta x^{\beta-1}\exp\left(-\alpha x^{\beta}\right)$ pour $x \geq 0$ et $\alpha, \beta > 0$.

a) Donnez la fonction de log-vraisemblance pour un échantillon aléatoire de N observations.

b) Donnez les équations de vraisemblance pour l'estimation par maximum de vraisemblance de α et β.

c) Obtenez la matrice des dérivées secondes de la log-vraisemblance par rapport à α et β. Les espérances exactes des éléments impliquant β impliquent des dérivées de la fonction Gamma. Ce qui est assez difficile à trouver analytiquement. Votre résultat exact donne un estimateur empirique. Comment estimeriez-vous plus facilement la matrice de covariance asymptotique de l'estimateur MV ?

d) Prouvez que $\alpha\beta Cov\left(\ln x\ , x^{\beta}\right) = 1$.
[Utilisez le fait que l'espérance des dérivées premières de la log-vraisemblance est nulle]

99. On veut tester les J hypothèses linéaires $R\beta = r$. On peut utiliser un test de Wald de l'estimateur des multiplicateurs de Lagrange des MCC, soit les hypothèses : $H_0 : \lambda = 0$. Le test de Wald est : $W = \lambda^{*\prime}\hat{V}(\lambda^*)^{-1}\lambda^*$.

a) Prouvez que ce test dépend du ratio des estimateurs de la variance de l'erreur :

$$W = (N-K)\left[\frac{e^{*\prime}e^*}{e'e} - 1\right]$$

b) **Pourquoi le facteur entre crochets est positif ?**

c) **Prouvez que cette statistique de test est égale à $J \times F$, où F est la statistique de test des restrictions $R\beta = r$ dans le modèle non contraint.**

d) **Montrez que la statistique de test du multiplicateur de Lagrange peut se réécrire comme**

$$LM = \frac{N \times J}{((N-K)/F) + J}$$

100. Dans un modèle *Logit* ou *Probit* où il n'y a que la constante : $y_i = \beta_1 + \varepsilon_i$. Notez P : le nombre d'observations avec une valeur $y = 1$ parmi les N observations.

a) **Donnez l'estimateur de *MV* de la constante dans le modèle *Logit* et *Probit*.**

b) **Quelle est la valeur maximale de la log-vraisemblance dans ces deux modèles ?**

c) **Proposez un test de significativité conjointe des paramètres de pente s'il y a d'autres régresseurs dans le modèle : $y_i = \beta_1 + \beta_2 x_2 + \cdots + \beta_K x_K + \varepsilon_i$.**