PROPRIETES STATISTIQUES DES MCO

Afin de connaître les propriétés statistiques des MCO on est conduit à faire des hypothèses sur les erreurs, ces hypothèses vont être la base des ennuis de l'économétrie : vérification ou non des hypothèses $\Longrightarrow$ entraîne construction de tests ; si l'hypothèses n'est pas vérifiée par quoi remplacer les MCO ?
On applique les MCO donc les hypothèses de base H_0 des MCO doivent être vérifiées

$$\begin{array}{lll} H_0^1 & : & n \geq k \\ H_0^2 & : & \text{la matrice } X \text{ doit être de plein rang k pour que } {}^tXX \text{ soit inversible} \\ & & \text{pas de relations linéaires entre les variables explicatives} \\ H_0^3 & : & \text{Les coefficients } a_i \text{ sont inconnus mais constants} \end{array}$$

1 Premières hypothèses et leurs conséquences

On donnera à ces hypothèses l'indice 1 ce sont les hypothèses probabilistes du modèle.

H_1^1 : les erreurs sont des variables aléatoires, les variables explicatives sont non aléatoires

H_1^2 : les erreurs ont une espérance nulle $E(\epsilon_t) = 0$ pour tout t

H_1^3 : les erreurs ont une variance constante inconnue notée σ^2, $\forall t \to V(\epsilon_t) = \sigma^2$ on parlera d'homoscédasticité

H_1^4 : les covariances des erreurs sont nulles $Cov(\epsilon_t, \epsilon_{t+h}) = 0$ Non Autocorrélation

Dans le modèle écrit matriciellement (voir MCO.PDF) $\overrightarrow{Y} = X\overrightarrow{a} + \overrightarrow{\epsilon}$, l'hypothèse H_1^1 indique que les variables explicatives ne sont pas aléatoires, or le vecteur $\overrightarrow{a}$ est inconnu mais constant donc $X\overrightarrow{a}$ est inconnu mais non aléatoire. Cette même hypothèse H_1^1 indique que le erreurs sont aléatoires donc le vecteur erreur $\overrightarrow{\epsilon}$ est un vecteur aléatoire et ainsi le vecteur $\overrightarrow{Y}$ qui est la somme d'un vecteur non aléatoire et d'un vecteur aléatoire devient lui-même aléatoire. Le résultat des MCO $\overrightarrow{\widehat{a}} = ({}^tXX)^{-1}\,{}^tX\,\overrightarrow{Y}$ est constitué par $({}^tXX)^{-1t}X$) qui ne dépendant que de X est non aléatoire et de $\overrightarrow{Y}$ qui est aléatoire, $\overrightarrow{\widehat{a}}$ est donc un vecteur aléatoire. L'estimateur des MCO est aléatoire.

$$\begin{aligned} \overrightarrow{\widehat{a}} &= ({}^tXX)^{-1}\,{}^tX\overrightarrow{Y} = ({}^tXX)^{-1}\,{}^tX(X\overrightarrow{a} + \overrightarrow{\epsilon}) \\ \overrightarrow{\widehat{a}} &= \overrightarrow{a} + ({}^tXX)^{-1}\,{}^tX\overrightarrow{\epsilon} \end{aligned}$$

1.1 Espérance mathématique de $\overrightarrow{\widehat{a}}$

Le vecteur $\overrightarrow{a}$ est inconnu mais non aléatoire de même que la matrice $({}^tXX)^{-1}\,{}^tX$ qui ne dépend que des variables explicatives. L'espérance du vecteur $\overrightarrow{\widehat{a}}$ est donc

$$\begin{aligned} E(\overrightarrow{\widehat{a}}) &= E(\overrightarrow{a}) + E(({}^tXX)^{-1}\,{}^tX\overrightarrow{\epsilon}) = \overrightarrow{a} + ({}^tXX)^{-1}\,{}^tXE(\overrightarrow{\epsilon}) \\ E(\overrightarrow{\widehat{a}}) &= \overrightarrow{a} + ({}^tXX)^{-1}\,{}^tX\,\overrightarrow{0} = \overrightarrow{a} \end{aligned}$$

Le vecteur des MCO est donc un estimateur sans biais du vecteur $\overrightarrow{a}$. Ce résultat est obtenu en utilisant seulement les hypothèses H_1^1 et H_1^2 donc quelques soient les propriétés des variances et covariances ce résultat reste vrai.

1.2 Matrice de variances-covariances de $\overrightarrow{\widehat{a}}$

Cette matrice a pour définition

$$V_{\overrightarrow{\widehat{a}}} = E[(\overrightarrow{\widehat{a}} - E(\overrightarrow{\widehat{a}}))\ {}^t(\overrightarrow{\widehat{a}} - E(\overrightarrow{\widehat{a}}))]$$

C'est une matrice carrée (k,k) symétrique. On montre quelle contient sur sa diagonale les variances des erreurs et hors diagonale les covariances.
On a vu que $E(\overrightarrow{\widehat{a}}) = \overrightarrow{a}$ et $\overrightarrow{\widehat{a}} - \overrightarrow{a} = ({}^tXX)^{-1t}X\overrightarrow{\epsilon}$ donc

$$\begin{aligned} V_{\overrightarrow{\widehat{a}}} &= E[(\overrightarrow{\widehat{a}} - \overrightarrow{a})\ {}^t(\overrightarrow{\widehat{a}} - \overrightarrow{a})] = E[(({}^tXX)^{-1}\ {}^tX\overrightarrow{\epsilon})({}^t\overrightarrow{\epsilon}X({}^tXX)^{-1})] \\ V_{\overrightarrow{\widehat{a}}} &= (({}^tXX)^{-1}\ {}^tX)E[\overrightarrow{\epsilon}\ {}^t\overrightarrow{\epsilon}](X({}^tXX)^{-1}) \end{aligned}$$

D'après la définition d'une matrice de variances-covariances, l'expression $E[\overrightarrow{\epsilon}\ {}^t\overrightarrow{\epsilon}]$ est la matrice de variances-covariances du vecteur $\overrightarrow{\epsilon}$ en utilisant l'hypothèse H_1^2 $E(\epsilon_t) = 0\ \forall t$. De plus l'hypothèse H_1^3 indique que les variances sont toutes égales à σ^2 et l'hypothèse H_1^4 que les covariances sont nulles. La matrice de variances-covariances des erreurs a donc σ^2 sur tous les éléments de sa diagonale et 0 ailleurs, c'est donc

$$V_{\overrightarrow{\epsilon}} = \sigma^2 I_n$$

où I_n est la matrice identité (n,n) avec des 1 sur la diagonale et 0 ailleurs.

$$\begin{aligned} V_{\overrightarrow{\widehat{a}}} &= (({}^tXX)^{-1}\ {}^tX)[\sigma^2 I_n](X({}^tXX)^{-1}) = \sigma^2(({}^tXX)^{-1}\ {}^tX)[I_n](X({}^tXX)^{-1}) \\ V_{\overrightarrow{\widehat{a}}} &= \sigma^2({}^tXX)^{-1} \end{aligned}$$

Pour ce résultat il faut que toutes les hypothèses soient vérifiées

1.3 Théorème de Gauss-Markov

A partir de maintenant, on se contentera des résultats (voir votre cours). On montre que l'estimateur des MCO est le meilleurs c'est-à dire à variance minimale pour ses composantes parmi tous les estimateurs linéaires et sans biais (la linéarité venant du fait que les coefficients du vecteur $\overrightarrow{\widehat{a}}$ sont des combinaisons linéaires des Y_t). On dit que l'estimateur des MCO est B.L.U.E.: best linear unbiaised estimator.

1.4 Estimateur de σ^2

On montre que l'estimateur sans biais de σ^2 que l'on notera s^2 est en notant SCR la somme des carrés des résidus

$$s^2 = \frac{\sum e_t^2}{(n-k)} = \frac{SCR}{(n-k)}$$

1.5 Estimateur de $V_{\overrightarrow{\widehat{a}}}$

$V_{\overrightarrow{\widehat{a}}} = \sigma^2({}^tXX)^{-1}$ est une matrice inconnue car les variables explicatives étant connues alors la matrice X est connue mais la constante σ^2 reste inconnue, on va donc l'estimer par son estimateur s^2 ce qui fournit une estimateur de $V_{\overrightarrow{\widehat{a}}}$

$$\widehat{V_{\overrightarrow{\widehat{a}}}} = s^2({}^tXX)^{-1}$$

Ce résultat permet de trouver les estimateurs des variances des coefficients des MCO: si on note h$_{ij}$ les éléments de la matrice $({}^tXX)^{-1}$ alors, les variances étant sur la diagonale, on aura la variance du coefficient $\widehat{a}_i$, $V(\widehat{a}_i) = \sigma^2 h_{ii}$ qui sera estimée par$\widehat{V(\widehat{a}_i)} = s^2 h_{ii}$. De même la covariance entre les coefficients $\widehat{a}_i$ et $\widehat{a}_j$, $Cov(\widehat{a}_i, \widehat{a}_j)$ sera estimé par $\widehat{Cov(\widehat{a}_i, \widehat{a}_j)} = s^2 h_{ij}$.

2 Asymptotique:

Si on ajoute à ces quatre hypothèse les hypothèses

H$_1^5$: si n$\to \infty$ alors $\lim\limits_{n\to\infty} \frac{{}^tXX}{n} = V_X$ matrice finie définie positive.

H$_1^6$: les erreurs sont indépendantes et suivent la même loi inconnue (i..i.d.)

- conséquence de H$_1^5$: la matrice de variances-covariances de $\overrightarrow{\widehat{a}}$ tend vers la matrice nulle (k,k) et l'estimateur des MCO est donc un estimateur convergent en moyenne quadratique de $\overrightarrow{a}$. En effet

  $$V_{\overrightarrow{\widehat{a}}} = \sigma^2({}^tXX)^{-1} = \frac{\sigma^2}{n}[\frac{1}{n}({}^tXX)]^{-1} \underset{n\to\infty}{\to} \frac{\sigma^2}{n} V_X^{-1} \to \underset{(k,k)}{0}$$

  Remarque : ce premier résultat aurait été valable également si on avait remplacé l'hypothèse H$_1^5$ par une hypothèse moins contraignante notée H$_1^5$ bis

  H$_1^5$ bis : si n$\to \infty$ alors $\lim\limits_{n\to\infty} {}^tXX = \underset{(k,k)}{0}$

  $$V_{\overrightarrow{\widehat{a}}} = \sigma^2({}^tXX)^{-1} \underset{n\to\infty}{\to} \sigma^2 \underset{(k,k)}{0}$$

- conséquence de H$_1^5$ et H$_1^6$: le théorème central limite montre que si n$\to \infty$ alors :

  $$\sqrt{n}(\overrightarrow{\widehat{a}} - \overrightarrow{a}) \to N_k(\overrightarrow{0}, \sigma^2 V_X^{-1})$$

3 Hypothèse de normalité des erreurs et conséquences

Si aux hypothèses H$_1^1$,H$_1^2$, H$_1^3$ et H$_1^4$ on ajoute l'hypothèse H$_1^7$

H$_1^1$: les erreurs sont des variables aléatoires, les variables explicatives sont non aléatoires

H$_1^2$: les erreurs ont une espérance nulle $E(\epsilon_t) = 0$ pour tout t

H$_1^3$: les erreurs ont une variance constante notée σ^2, $\forall t \to V(\epsilon_t) = \sigma^2$ Homoscédasticité

H$_1^4$: les covariances des erreurs sont nulles $Cov(\epsilon_t, \epsilon_{t+h}) = 0$ Non Autocorrélation

H_1^7 : les erreurs suivent la même loi Normale $N(0, \sigma^2)$

3.1 Indépendance des erreurs

Si l'hypothèse H_1^7 est vérifiée en plus des 4 premières alors les erreurs qui n'étaient pas corrélées sont donc indépendantes. De plus les lois sont i.i.d.

3.2 Autre propriété de l'estimateur des MCO

On montre que l'estimateur des MCO est aussi l'estimateur du maximum de vraisemblance de $\overrightarrow{a}$ qui est toujours sans biais. Par contre on montre que l'estimateur du maximum de vraisemblance de σ^2 est $\widehat{\sigma^2} = \frac{\sum e_t^2}{n} = \frac{SCR}{n}$ qui n'est qu'asymptotiquement sans biais. Le meilleur estimateur de σ^2 reste toujours s$^2 = \frac{SCR}{n}$

3.3 Lois de probabilité conséquences de H$_1^7$

Si chaque erreur ϵ_t suit une loi Normale $N(0, \sigma^2)$, alors le vecteur erreur $\overrightarrow{\epsilon}$ suit une loi Normale à n dimensions

$$\overrightarrow{\epsilon} \text{ suit } N_n(\overrightarrow{0}, \sigma^2 I_n)$$

3.3.1 Théorème 1

Dans la suite on va utiliser le théorème suivant sur la loi suivie par un vecteur défini comme transformé d'un vecteur Normal.
Soit un vecteur aléatoire $\underset{(m,1)}{\overrightarrow{Z}}$ suivant une loi Normale $N_m(E(\overrightarrow{Z}), V_{\overrightarrow{Z}})$. On note $\underset{(m',1)}{\overrightarrow{W}}$ le vecteur transformé de $\overrightarrow{Z}$ à l'aide de la matrice $\underset{(m',m)}{A}$.

$$\overrightarrow{W} = A\overrightarrow{Z}$$

avec $m\prime \leq m$.

- Si la matrice A est de plein rang $m\prime$ (car $m\prime \leq m$) alors le vecteur $\overrightarrow{W}$ suit aussi une loi Normale $N_{m'}[E(\overrightarrow{W}) = AE(\overrightarrow{Z}); V_{\overrightarrow{W}} = AV_{\overrightarrow{Z}}\ {}^tA]$ et chaque composante de $\overrightarrow{W}$ suit une loi Normale à une dimension.

- Si la matrice A n'est pas de plein rang (de rang r<m') alors le vecteur $\overrightarrow{W}$ ne suit plus une loi Normale à m' , mais une loi dite dégénérée de rang seulement r. Chaque composante de $\overrightarrow{W}$ ne suit plus une loi Normale à une dimension.

3.3.2 loi de $\overrightarrow{Y}$

Le vecteur $\overrightarrow{Y} = X\overrightarrow{a} + \overrightarrow{\epsilon}$ a pour espérance mathématique $E(\overrightarrow{Y}) = E(X\overrightarrow{a}) + E(\overrightarrow{\epsilon})$, or $X\overrightarrow{a}$ est non aléatoire et $E(\overrightarrow{\epsilon}) = \overrightarrow{0}$ donc $E(\overrightarrow{Y}) = X\overrightarrow{a}$.
La matrice de variances-covariances de $\overrightarrow{Y}$ a pour définition $V_{\overrightarrow{Y}} = E[(\overrightarrow{Y} - E(\overrightarrow{Y}))\ {}^t(\overrightarrow{Y} - E(\overrightarrow{Y}))] = E[(\overrightarrow{Y} - X\overrightarrow{a})\ {}^t(\overrightarrow{Y} - X\overrightarrow{a})] = E[\overrightarrow{\epsilon}\ {}^t\overrightarrow{\epsilon}] = \sigma^2 I_n$
D'après le théorème 1 ci-dessus $\overrightarrow{Y} - X\overrightarrow{a} = \overrightarrow{\epsilon} = I_n\overrightarrow{\epsilon}$ donc on passe du vecteur $\overrightarrow{\epsilon}$ au vecteur $\overrightarrow{Y - X a}$ par la matrice I_n qui est de plein rang n.

$$\overrightarrow{Y} \text{ suit } N_n(X\overrightarrow{a}, \sigma^2 I_n)$$

3.3.3 loi de $\overrightarrow{\widehat{Y}}$

Le vecteur estimateur de $\overrightarrow{Y}$ est $\overrightarrow{\widehat{Y}} = X\overrightarrow{\widehat{a}} = X({}^tXX)^{-1}\,{}^tX\overrightarrow{Y} = X({}^tXX)^{-1}\,{}^tX(X\overrightarrow{a} + \overrightarrow{\epsilon})$ en remplaçant $\overrightarrow{Y}$ par sa valeur en fonction de $\overrightarrow{\epsilon}$
$\overrightarrow{\widehat{Y}} = X({}^tXX)^{-1}\,{}^tXX\overrightarrow{a} + X({}^tXX)^{-1t}X\overrightarrow{\epsilon} = X\overrightarrow{a} + N\overrightarrow{\epsilon}$ en posant $N = X({}^tXX)^{-1t}X$, la matrice X étant de dimension (n,k) , la matrice N est également (n,k) donc de rang k>n . On peut remarquer que $N =^t N = N^2$ c'est une matrice idempotente de projection orthogonale de $\overrightarrow{Y}$ sur la variété linéaire H_k(résultat qui est la base des MCO, voir le chapitre 2). L'espérance mathématique de $\overrightarrow{\widehat{Y}}$ est $E(\overrightarrow{\widehat{Y}}) = X\overrightarrow{a}$ car N est une matrice non aléatoire et $E(\overrightarrow{\epsilon}) = \overrightarrow{0}$. La matrice de variances-covariances s'écrit
$V_{\overrightarrow{\widehat{Y}}} = E[(\overrightarrow{\widehat{Y}} - E(\overrightarrow{\widehat{Y}}))\ {}^t(\overrightarrow{\widehat{Y}} - E(\overrightarrow{\widehat{Y}}))] = E[(\overrightarrow{\widehat{Y}} - X\overrightarrow{a})\ {}^t(\overrightarrow{\widehat{Y}} - X\overrightarrow{a})] = E[N\overrightarrow{\epsilon}\ {}^t\overrightarrow{\epsilon}\,{}^tN]$
$V_{\overrightarrow{\widehat{Y}}} = NE[\overrightarrow{\epsilon}\ {}^t\overrightarrow{\epsilon}]^tN = \sigma^2 N^tN = \sigma^2 N^2 = \sigma^2 N$
D'après le théorème 1: $\overrightarrow{\widehat{Y}} - X\overrightarrow{a} = N\overrightarrow{\epsilon}$ on passe donc du vecteur $\overrightarrow{\epsilon}$ au vecteur centré $\overrightarrow{\widehat{Y}} - X\overrightarrow{a}$ à l'aide de la matrice de transformation N qui comme on l'a vu n'est pas de plein rang, le vecteur $\overrightarrow{\widehat{Y}} - X\overrightarrow{a}$ ne suit plus une loi Normale à n dimensions , mais une loi dite dégénérée de rang seulement k. Chaque composante de $\overrightarrow{\widehat{Y}} - X\overrightarrow{a}$ ne suit plus une loi Normale à une dimension.

3.3.4 loi du vecteur des résidus

Le vecteur $\overrightarrow{e}$ des résidus a pour définition $\overrightarrow{e} = \overrightarrow{Y} - \overrightarrow{\widehat{Y}} = X\overrightarrow{a} + \overrightarrow{\epsilon} - X\overrightarrow{\widehat{a}})$
or $\overrightarrow{\widehat{a}} = \overrightarrow{a} + ({}^tXX)^{-1}\,{}^tX\overrightarrow{\epsilon}$ donc

$$\begin{aligned}
\overrightarrow{e} &= X\overrightarrow{a} + \overrightarrow{\epsilon} - X(\overrightarrow{a} + ({}^tXX)^{-1}\,{}^tX\overrightarrow{\epsilon}) \\
\overrightarrow{e} &= \overrightarrow{\epsilon} - X({}^tXX)^{-1}\,{}^tX\overrightarrow{\epsilon} = (I_n - X({}^tXX)^{-1}\,{}^tX)\overrightarrow{\epsilon} = M\overrightarrow{\epsilon}
\end{aligned}$$

où la matrice $\underset{(n,n)}{M}$ est telle que comme on peut le vérifier $M =^t M = M^2$, par construction le vecteur résidu est orthogonal à la variété linéaire H_k, la matrice M est donc la matrice de projection sur H_{n-k} , complémentaire de H_k dans R^n. On remarque $M = I_n - N$ soit $M + N = I_n$, les matrices définissent bien des projections orthogonales sur des espaces complémentaires. La matrice M de dimension (n,n)est de rang (n-k).
On a $E(\overrightarrow{e}) = E(M\overrightarrow{\epsilon}) = ME(\overrightarrow{\epsilon}) = \overrightarrow{0}$
$V_{\overrightarrow{e}} = E(\overrightarrow{e}\ {}^t\overrightarrow{e}) = E(M\overrightarrow{\epsilon}\ {}^t\overrightarrow{\epsilon}\,{}^tM) = ME(\overrightarrow{\epsilon}\ {}^t\overrightarrow{\epsilon})^tM = M\sigma^2 I_n\ {}^tM = \sigma^2 M^tM =$
$V_{\overrightarrow{e}} = \sigma^2 M^2 = \sigma^2 M$
D'après le théorème 1: $\overrightarrow{e} = M\overrightarrow{\epsilon}$ on passe donc du vecteur $\overrightarrow{\epsilon}$ au vecteur centré $\overrightarrow{e}$ à l'aide de la matrice de transformation M qui comme on l'a vu n'est pas de plein rang, le vecteur $\overrightarrow{e}$ ne suit plus une loi Normale à n dimensions , mais une loi dite dégénérée de rang seulement n-k. Chaque composante de $\overrightarrow{e}$ ne suit plus une loi Normale à une dimension.

3.3.5 Théorème 2

Théorème 2 : Si $\overrightarrow{\epsilon}$ suit une loi Normale à n dimensions (Hypothèse H_7) alors le scalaire ${}^t\overrightarrow{\epsilon} M \overrightarrow{\epsilon} /\ \sigma^2$ suit une loi du χ^2_{n-k} à n-k degrés de liberté.
Démonstration:
On part du fait que la matrice M (de dimension (n,n) et de rang (n-k) est symétrique. On peut donc la décomposer à l'aide
- d'une matrice V de plein rang n, orthogonale telle que ${}^tVV$=I_n
- et de sa matrice des valeurs propres Λ matrice diagonale qui contient les valeurs propres de M. Comme M est une matrice de projection orthogonale sur H_{n-k}, ses valeurs propres sont des 1 ou des 0 avec (n-k) valeurs propres égales à 1 et k égales à 0.

$$\begin{aligned} M &= V\Lambda^t V \\ {}^t\overrightarrow{\epsilon} M \overrightarrow{\epsilon} &= {}^t\overrightarrow{\epsilon} V\Lambda^t V \overrightarrow{\epsilon} \end{aligned}$$

On pose $\overrightarrow{u} = {}^tV\overrightarrow{\epsilon}$
on a E($\overrightarrow{u}$) = ${}^tVE(\overrightarrow{\epsilon})$ et $V_{\overrightarrow{u}} = E({}^tV\overrightarrow{\epsilon}\,{}^t\overrightarrow{\epsilon}V) = {}^tVE(\overrightarrow{\epsilon}\,{}^t\overrightarrow{\epsilon})V = {}^tV\sigma^2 I_n V = \sigma^2\ {}^tVV = \sigma^2 I_n$
Si $\overrightarrow{\epsilon}$ suit une loi normale à n dimensions $N_n(\overrightarrow{0}, \sigma^2 I_n)$, comme V est de plein rang chaque u_t suit une loi N(0,σ^2) donc chaque u_t^2/σ^2 suit un χ^2_1

$${}^t\overrightarrow{\epsilon} M \overrightarrow{\epsilon} =^t \overrightarrow{\epsilon} V\Lambda^t V \overrightarrow{\epsilon} =^t \overrightarrow{u}\Lambda\overrightarrow{u} = \sum_1^n u_t^2\lambda_t = \sum_1^{n-k} u_t^2\lambda_t$$

car k valeurs de λ_t sont nulles.

$$\frac{{}^t\overrightarrow{\epsilon} N \overrightarrow{\epsilon}}{\sigma^2} = \frac{\sum_1^{n-k} u_t^2\lambda_t}{\sigma^2} \text{ suit donc un } \chi^2_{n-k}$$

3.3.6 loi de s^2

Application du théorème 2:
On a vu que $\overrightarrow{e} = M\overrightarrow{\epsilon}$ donc ${}^t\overrightarrow{e}\,\overrightarrow{e} = {}^t\overrightarrow{\epsilon}\ {}^tMM\overrightarrow{\epsilon} = {}^t\overrightarrow{\epsilon}\ M\overrightarrow{\epsilon}$ car M est une matrice de projection orthogonale.

$$\frac{{}^t\overrightarrow{e}\,\overrightarrow{e}}{\sigma^2} = \frac{\sum_1^n e_t^2}{\sigma^2} = \frac{SCR}{\sigma^2} \text{ suit la loi du } \chi^2_{n-k}$$

or $s^2 = SCR/(n-k) \implies SCR = (n-k)s^2$ donc

$$\frac{(n-k)s^2}{\sigma^2} \text{ suit la loi du } \chi^2_{n-k}$$