ESTIMATION D'UN MODELE DE REGRESSION LINEAIRE PAR LA METHODE DU MAXIMUM DE VRAISEMBLANCE

A côté de la méthode des moindres carrés ordinaires (MCO), la méthode du maximum de vraisemblance/MV (en anglais « Maximum Likelihood method » ou ML method) permet aussi d'estimer les paramètres d'un modèle de régression, sous l'hypothèse que la vraie (loi) distribution desdits paramètres est connue(1). Si le principe pour les MCO est de trouver le paramètre qui minimise la somme des carrés des erreurs, la méthode du maximum de vraisemblance cherche par contre à trouver le paramètre à même (ayant une forte probabilité) de reproduire les vraies valeurs de l'échantillon (celles réellement observées), soit trouver la valeur la plus *vraisemblable* du paramètre d'une population partant d'un échantillon donné (lire Bosonga Bofeki L. JP., 2019, p. 131). Autrement dit, pour reprendre les propos (de la même veine) chers à Kintambu Mafuku, « *la méthode du MV est basée sur l'idée que si nous nous trouvons en présence des possibles valeurs différentes pour un paramètre, nous choisirons la valeur avec laquelle le modèle génèrerait avec plus de probabilité l'échantillon observé* » (Kintambu Mafuku E.G., 2004, p. 76).

L'on notera aussi que, sous l'hypothèse que les erreurs sont normalement distribuées, les estimateurs des MCO et ceux du maximum de vraisemblance sont identiques comme démontré plus bas.

Par ailleurs, il tient de préciser que l'estimateur de maximum de vraisemblance sert de base à certains tests statistiques, notamment : le test de Wald, celui du ratio de vraisemblance et celui du multiplicateur de Lagrange.

Dans les lignes qui suivent, nous montrons comment estimer les paramètres d'un modèle de régression linéaire simple, autant pour un modèle de régression multiple, par la méthode du maximum de vraisemblance ; ensuite, nous présentons les trois tests d'hypothèses construits sur base de l'estimateur du maximum de vraisemblance.

1. *Estimateur du maximum de vraisemblance (ML)*

a) *Estimateur ML d'un modèle de régression linéaire simple (MRLS)*

Considérons le MRLS suivant :

$$Y_t = a_0 + a_1 X_t + u_t \ldots [3.11]$$

Y_t, linéairement dépendant du terme d'erreur « u_t », est une variable normalement distribuée de paramètres (moyenne et variance) : $E(Y_t) = a_0 + a_1 X_t$ et $V(Y_t) = \sigma_u^2$.

En effet, si « $u_t \sim N(0, \sigma^2) \rightarrow E(u_t) = 0$ », alors : $E(Y_t) = E(a_0 + a_1 X_t + u_t) = a_0 + a_1 X_t$.

Puisque : $Y_t - E(Y_t) = (a_0 + a_1 X_t + u_t) - (a_0 + a_1 X_t) = u_t$, alors la variance de Y_t est :

$V(Y_t) = E[Y_t - E(Y_t)]^2 = E(u_t)^2 = \sigma_u^2$.

D'où, la distribution de Y_t : $Y_t \sim N[(a_0 + a_1 X_t); \sigma_u^2]$.

[1] L'application de la méthode du maximum de vraisemblance exige de connaitre a priori la loi de distribution des paramètres, contrairement à la méthode des MCO qui permet d'estimer les paramètres sans avoir à connaitre leur distribution a priori.

Fonction de densité de probabilité jointe

Elle s'écrit :$f(Y_1, Y_2, \dots, Y_t | a_0 + a_1 X_t, \sigma_u^2)$. Si les « Y_t » sont indépendantes, cette fonction peut s'écrire aussi :

$$f(Y_1, Y_2, \dots, Y_t | a_0 + a_1 X_t, \sigma_u^2) = f(Y_1 | a_0 + a_1 X_t, \sigma_u^2) * f(Y_2 | a_0 + a_1 X_t, \sigma_u^2) * \dots * f(Y_t | a_0 + a_1 X_t, \sigma_u^2) \dots [3.11a]$$

Fonction de densité de la loi normale générale

En général, cette fonction se présente comme suit ($a_0 \; et \; a_1$ sont considérés comme des coefficients estimés) :

$$f(Y_t) = \frac{1}{\sqrt{2\pi}\sigma_u} \exp\left[-\frac{1}{2}\left(\frac{Y_t - a_0 - a_1 X_t}{\sigma_u}\right)^2\right] \dots [3.11b]$$

Fonction de vraisemblance et fonction log-vraisemblance

______*Fonction de vraisemblance*

Elle est obtenue lorsqu'on remplace la fonction de densité de la loi normale [3.11b] dans la fonction de densité de probabilité jointe [3.11a], en supposant connues les « $Y_1, Y_2, \dots, Y_t$ » :

$$f(Y_1, Y_2, \dots, Y_t | a_0 + a_1 X_t, \sigma_u^2) = \frac{1}{\left(\sqrt{2\pi}\right)^t \sigma_u^t} \exp\left[-\frac{1}{2}\sum\left(\frac{Y_t - a_0 - a_1 X_t}{\sigma_u}\right)^2\right]$$

$$ou \; encore : \; L(a_0, a_1, \sigma_u^2) = \frac{1}{\left(\sqrt{2\pi}\right)^t \sigma_u^t} \exp\left[-\frac{1}{2}\sum\left(\frac{Y_t - a_0 - a_1 X_t}{\sigma_u}\right)^2\right] \dots [3.11c]$$

______*Fonction log-vraisemblance*

Lorsqu'on effectue une transformation logarithmique de la fonction de vraisemblance ci-dessus, l'on obtient la fonction dite log-vraisemblance qui servira de base à l'estimation des paramètres « $\hat{a}_0 \; et \; \hat{a}_1$ ». La fonction log-vraisemblance[2] s'écrit :

$$\ln L = \ln\left(\frac{1}{\left(\sqrt{2\pi}\right)^t \sigma_u^t}\right) + \ln\left(\exp\left[-\frac{1}{2}\sum\left(\frac{Y_t - a_0 - a_1 X_t}{\sigma_u}\right)^2\right]\right)$$

En effet : $\ln 1 = 0$; $\ln \exp(a) = \ln e^a = a$; $\ln x^2 = 2 \ln x$; $\sqrt{2\pi} = (2\pi)^{1/2}$.

$$\ln\left(\frac{1}{\left(\sqrt{2\pi}\right)^t \sigma_u^t}\right) = \ln 1 - \left[\ln\left(\sqrt{2\pi}\right)^t + \ln \sigma_u^t\right] = -\frac{t}{2}\ln(2\pi) - t \ln \sigma_u$$

$$\ln\left(\exp\left[-\frac{1}{2}\sum\left(\frac{Y_t - a_0 - a_1 X_t}{\sigma_u}\right)^2\right]\right) = -\frac{1}{2}\sum\left(\frac{Y_t - a_0 - a_1 X_t}{\sigma_u}\right)^2$$

$$\rightarrow \ln L = -\, t \ln \sigma_u - \frac{t}{2}\ln(2\pi) - \frac{1}{2}\sum\left(\frac{Y_t - a_0 - a_1 X_t}{\sigma_u}\right)^2$$

[2] La fonction « $\ln L$ » est similaire à « L », c'est dire qu'on peut prendre l'une ou l'autre, car le maximum de « $\ln L$ » est atteint au même point que « L » (L étant une fonction monotone). L'on préfère maximiser « $\ln L$ » pour estimer les paramètres inconnus vu que cette fonction simplifie les calculs.

Considérant que : $\ln \sigma_u^2 = 2 \ln \sigma_u \rightarrow \frac{1}{2} \ln \sigma_u^2 = \ln \sigma_u$, alors l'on peut écrire (*fonction log-vraisemblance retenue pour l'estimation des paramètres*) :

$$\rightarrow \ln L = -\frac{t}{2} \ln \sigma_u^2 - \frac{t}{2} \ln(2\pi) - \frac{1}{2} \sum \left(\frac{Y_t - a_0 - a_1 X_t}{\sigma_u} \right)^2 \dots [3.11d]$$

Estimation des paramètres « $\hat{a}_0, \hat{a}_1 et\ \sigma_u^2$ »

Pour estimer les paramètres « $\hat{a}_0, \hat{a}_1, \hat{\sigma}_u^2$ » par le maximum de vraisemblance, la démarche va consister à maximiser la fonction log-vraisemblance ci-dessus $[3.11d]$, ce qui revient à annuler ses dérivées premières par rapport aux arguments « $a_0,\ a_1\ et\ \sigma_u^2$ » comme suit :

En effet : $\left(\left[G[Y(x)] \right]^n \right)' = n\ G^{n-1}\ Y(x)'$

$$\frac{\partial \ln L}{\partial a_0} = -\frac{1}{2} * 2 * \sum \left(\frac{Y_t - a_0 - a_1 X_t}{\sigma_u} \right)^{2-1} (-1) = -\sum \left(\frac{Y_t - a_0 - a_1 X_t}{\sigma_u} \right) (-1) = 0 \dots (a)$$

$$\frac{\partial \ln L}{\partial a_1} = -\frac{1}{2} * 2 * \sum \left(\frac{Y_t - a_0 - a_1 X_t}{\sigma_u} \right)^{2-1} (-X_t) = -\sum \left(\frac{Y_t - a_0 - a_1 X_t}{\sigma_u} \right) (-X_t) = 0 \dots (b)$$

$$\frac{\partial \ln L}{\partial \sigma_u^2} = \frac{\partial}{\partial \sigma_u^2} \left(-\frac{t}{2} \ln \sigma_u^2 \right) - \frac{1}{2} * \frac{\partial}{\partial \sigma_u^2} \left(\sum \left(\frac{Y_t - a_0 - a_1 X_t}{\sigma_u} \right)^2 \right) = 0 \dots (c)$$

On a :

$$\frac{\partial}{\partial \sigma_u^2} \left(-\frac{t}{2} \ln \sigma_u^2 \right) = -\frac{t}{2} * \frac{1}{\sigma_u^2} = -\frac{t}{2\sigma_u^2} \dots (d)$$

Appelons : $\sum \left(\frac{Y_t - a_0 - a_1 X_t}{\sigma_u} \right) = \sum(\bullet)$

$$\frac{\partial}{\partial \sigma_u^2} \left[\left(\sum(\bullet) \right)^2 \right] = \frac{\partial}{\partial \sigma_u} \left(\frac{\partial}{\partial \sigma_u} \left[\left(\sum(\bullet) \right)^2 \right] \right)$$

$$\frac{\partial}{\partial \sigma_u} \left[\left(\sum(\bullet) \right)^2 \right] = \sum [\ 2 * [(Y_t - a_0 - a_1 X_t) * \sigma_u^{-1}]^{\ 2-1} * (-1)(Y_t - a_0 - a_1 X_t) * \sigma_u^{-1-1}]$$

$$\rightarrow \frac{\partial}{\partial \sigma_u} \left[\left(\sum(\bullet) \right)^2 \right] = \sum -2 * \left(\frac{Y_t - a_0 - a_1 X_t}{\sigma_u} \right) * (Y_t - a_0 - a_1 X_t) * \sigma_u^{-2}$$

$$\frac{\partial}{\partial \sigma_u} \left(\frac{\partial}{\partial \sigma_u} \left[\left(\sum(\bullet) \right)^2 \right] \right) = \sum -2 * (-2) * \frac{(Y_t - a_0 - a_1 X_t)^2}{\sigma_u * \sigma_u^3} = 4 \sum \left[\frac{Y_t - a_0 - a_1 X_t}{\sigma_u^2} \right]^2 \dots (e)$$

Remplaçons les expressions (e) et (d) dans (c), l'on a :

$$\frac{\partial \ln L}{\partial \sigma_u^2} = -\frac{t}{2\sigma_u^2} + \frac{1}{2} \sum \left(\frac{Y_t - a_0 - a_1 X_t}{\sigma_u^2} \right)^2 = 0 \dots (c)^*$$

Si l'on développe cette expression, l'on obtient :

$$\frac{\partial \ln L}{\partial \sigma_u^2} = -\frac{t}{2\sigma_u^2} + \frac{1}{2} * \frac{1}{\sigma_u^4} \sum (Y_t - a_0 - a_1 X_t)^2 = -t + \frac{2\sigma_u^2}{2\sigma_u^4} \sum (Y_t - a_0 - a_1 X_t)^2 = 0$$

$$\rightarrow \frac{\partial \ln L}{\partial \sigma_u^2} = -t + \frac{1}{\sigma_u^2} \sum (Y_t - a_0 - a_1 X_t)^2 = 0 \dots (c)^{**}$$

Egalisées à zéro, les expressions (a), (b) et $(c)^{**}$ s'écrivent (*avec :* $\hat{a}_{0MV}$ = *estimateur du maximum de vraisemblance de* a_0) :

$$\frac{\partial \ln L}{\partial a_0} = 0 \rightarrow \sum (Y_t - \hat{a}_{0MV} - \hat{a}_{1MV}X_t) = 0 \rightarrow \sum Y_t = T\hat{a}_{0MV} + \hat{a}_{1MV}\sum X_t$$

$$\frac{\partial \ln L}{\partial a_1} = 0 \rightarrow \sum (Y_t - \hat{a}_{0MV} - \hat{a}_{1MV}X_t)\,X_t = 0 \rightarrow \sum X_tY_t = \hat{a}_{0MV}\sum X_t + \hat{a}_{1MV}\sum X_t^2$$

$$\frac{\partial \ln L}{\partial \sigma_u^2} = 0 \rightarrow -t + \frac{1}{\sigma_{uMV}^2}\sum (Y_t - \hat{a}_{0MV} - \hat{a}_{1MV}X_t)^2 = 0 \rightarrow \sigma_{uMV}^2 = \frac{1}{T}\sum (Y_t - \hat{a}_{0MV} - \hat{a}_{1MV}X_t)^2$$

Les deux premières expressions étant identiques aux équations normales fournies par les MCO, l'on déduit que les estimateurs MCO des paramètres « a_0 *et* a_1 » et les estimateurs du maximum de vraisemblance sont égaux ou les mêmes : $\hat{a}_{0MV} = \hat{a}_{0MCO}$ *et* $\hat{a}_{1MV} = \hat{a}_{1MCO}$. Par contre, le développement de la dernière expression donne un estimateur du maximum de vraisemblance « $\hat{\sigma}_{uMV}^2$ » de la variance de l'erreur « σ_u^2 » différent de l'estimateur MCO :

$$\hat{\boldsymbol{\sigma}}_{\boldsymbol{u}\,\boldsymbol{MV}}^{\boldsymbol{2}} = \frac{1}{T}\sum (Y_t - \hat{a}_0 - \hat{a}_1X_t)^2 = \frac{\boldsymbol{1}}{\boldsymbol{T}}\sum \boldsymbol{e_t^2} \neq \hat{\boldsymbol{\sigma}}_{\boldsymbol{u}\,\boldsymbol{MCO}}^{\boldsymbol{2}} = \frac{\boldsymbol{1}}{(\boldsymbol{T-2})}\sum e_t^2$$

Notons que, l'estimateur MCO de la variance des erreurs étant sans biais, l'estimateur du maximum de vraisemblance est biaisé, mais il reste convergent. Cette dernière propriété garantit la minimisation du biais avec l'accroissement de la taille de l'échantillon.

b) Estimateur ML d'un modèle de régression linéaire multiple (MRLM)

<u>*Forme fonctionnelle d'un MRLM*</u>

Un modèle de régression linéaire multiple ou modèle linéaire général, soit une généralisation de la régression simple au cas multivarié (où on a k variables explicatives, avec $k > 1$), s'écrit :

$$Y_t = a_0 + a_1X_{1t} + a_2X_{2t} + \cdots + a_iX_{it} + \cdots + a_kX_{kt} + u_t \ldots [3.11e]$$

Ou encore (sans constante) :

$$Y_t = a_1X_{1t} + a_2X_{2t} + \cdots + a_iX_{it} + \cdots + a_kX_{kt} + u_t \ldots [3.11e]^*$$

Avec : $t = 1, \ldots, T$ *(les observations) ;* Y_t = *la variable dépendante observée au temps t ;* $X_{1t}, \ldots, X_{kt}$ = *les k variables explicatives ou « régresseurs » ;* a_0 = *un paramètre du modèle ou terme constant (constante) ;* $a_1, \ldots, a_k$ = *les paramètres réels et inconnus du modèle ;* u_t = *le terme d'erreur.*

Sous forme matricielle, la relation $[3.11e]$ peut encore s'écrire comme suit :

$$\underbrace{Y}_{(T,1)} = \underbrace{X}_{(T,k+1)}\underbrace{a}_{(k+1,1)} + \underbrace{u}_{(T,1)} \rightarrow Y = Xa + u \ldots [3.11f]$$

Avec : $Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_t \\ \vdots \\ Y_T \end{pmatrix}$; $X = \begin{pmatrix} 1 & X_{11} & X_{21} & \cdots & X_{k1} \\ 1 & X_{12} & X_{22} & \cdots & X_{k2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & X_{1t} & X_{2t} & \cdots & X_{kt} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & X_{1T} & X_{2T} & \cdots & X_{kT} \end{pmatrix}$; $a = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_k \end{pmatrix}$; et $u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_t \\ \vdots \\ u_T \end{pmatrix}$.

Les « 1 » sur la première colonne de la matrice X captent les constantes « a_0 » dans le modèle.

Par contre, la notation matricielle de la relation sans constante $[3.11e]^*$ est (les formats de X et a changent) :

$$\underbrace{Y}_{(T,1)} = \underbrace{X}_{(T,k)} \underbrace{a}_{(k,1)} + \underbrace{u}_{(T,1)} \rightarrow Y = Xa + u \ldots [3.11f]^*$$

Y_t, linéairement dépendant du terme d'erreur « u » (*les erreurs sont supposées indépendantes et normalement distribuées, avec une espérance nulle ou $E(u) = 0$ et une variance constante* « σ_u^2 »), est une variable normalement distribuée de paramètre (moyenne et variance) :

$$E(Y_t) = a_0 + a_1 X_{1t} + a_2 X_{2t} + \cdots + a_k X_{kt} = Xa \ldots\ldots [3.11f]^{**}$$

$$V(Y_t) = \sigma_u^2 \ldots\ldots [3.11f]^{***}$$

<u>Densité de probabilité de Y_t</u>

La densité de probabilité de Y_t, connaissant les paramètres « a et σ_u^2 », est :

$$f(Y_t|a, \sigma_u^2) = \frac{1}{\sqrt{2\pi\sigma_u^2}} \exp\left[-\frac{1}{2\sigma_u^2}(Y_t - X_t a)^2\right] \ldots\ldots [3.11g]$$

<u>Fonction de densité de probabilité jointe ou conjointe de $Y_1, Y_2, \ldots, Y_t$</u>

Connaissant les paramètres « a et σ_u^2 », cette fonction s'écrit :$f(Y_1, Y_2, \ldots, Y_t|a, \sigma_u^2)$. Si les « Y_t » sont indépendantes, cette fonction(3) peut s'écrire aussi :

$$f(Y_1, Y_2, \ldots, Y_t|a, \sigma_u^2) = f(Y_1|a, \sigma_u^2) * f(Y_2|a, \sigma_u^2) * \ldots * f(Y_t|a, \sigma_u^2) = \prod_{t=1}^{T} f(Y_t|a, \sigma_u^2)$$

$$\rightarrow \prod_{t=1}^{T} f(Y_t|a, \sigma_u^2) = \left(\frac{1}{\sqrt{2\pi\sigma_u^2}}\right)^T \exp\left[-\frac{1}{2\sigma_u^2}\sum_{t=1}^{T}(Y_t - X_t a)^2\right]$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma_u^2}}\right)^T \exp\left[-\frac{1}{2\sigma_u^2}(Y - Xa)'(Y - Xa)\right] \ldots\ldots [3.11g]^a$$

<u>Fonction de vraisemblance et fonction log-vraisemblance</u>

______*Fonction de vraisemblance*

La *fonction de vraisemblance, notée « L »,* correspond à la relation « $[3.11g]^a$ » en supposant connus les « $Y_1, Y_2, \ldots, Y_t$ » et inconnus les paramètres « a et σ_u^2 » (à trouver), ce qui revient à écrire :

$$L(a, \sigma_u^2) = f(a, \sigma_u^2|Y_1, Y_2, \ldots, Y_t) = \left(\frac{1}{\sqrt{2\pi\sigma_u^2}}\right)^T \exp\left[-\frac{1}{2\sigma_u^2}(Y - Xa)'(Y - Xa)\right] \ldots\ldots [3.11g]^b$$

______*Fonction log-vraisemblance*

La transformation logarithmique de la fonction de vraisemblance ci-dessus donne la fonction dite « log-vraisemblance » qui servira de base à l'estimation des paramètres *« a »*. La fonction log-vraisemblance s'écrit :

$$\ln L = \ln\left[\left(\frac{1}{\sqrt{2\pi\sigma_u^2}}\right)^T\right] + \ln\left(\exp\left[-\frac{1}{2\sigma_u^2}(Y - Xa)'(Y - Xa)\right]\right) \ldots\ldots [3.11g]^c$$

[3] La fonction de densité de probabilité conjointe des Y_t correspond à la probabilité d'observer conjointement les Y_t.

En effet : $\ln(uv) = \ln u + \ln v$; $\ln \exp(a) = \ln e^{a} = a$; $\ln x^2 = 2\ln x$; $1/\sqrt{2\pi\sigma_u^2} = (2\pi\sigma_u^2)^{-1/2}$.

$$\rightarrow \mathbf{ln}\left[\left(\frac{\mathbf{1}}{\sqrt{\mathbf{2}\boldsymbol{\pi}\boldsymbol{\sigma}_{\boldsymbol{u}}^{\mathbf{2}}}}\right)^{\boldsymbol{T}}\right] = \ln\left[(2\pi\sigma_u^2)^{-\frac{T}{2}}\right] = -\frac{T}{2}[\ln\pi + \ln(2\sigma_u^2)] = -\frac{T}{2}\ln\pi - \frac{T}{2}\ln(2\sigma_u^2)$$

$$\rightarrow \mathbf{ln}\left(\mathbf{exp}\left[-\frac{\mathbf{1}}{\mathbf{2}\boldsymbol{\sigma}_{\boldsymbol{u}}^{\mathbf{2}}}(\boldsymbol{Y}-\boldsymbol{Xa})'(\boldsymbol{Y}-\boldsymbol{Xa})\right]\right) = -\frac{1}{2\sigma_u^2}(Y-Xa)'(Y-Xa)$$

$$\rightarrow (\boldsymbol{Y}-\boldsymbol{Xa})'(\boldsymbol{Y}-\boldsymbol{Xa}) = Y'Y - Y'Xa - X'a'Y + X'Xa'a = Y'Y - 2X'aY + X'Xa^2$$

Avec : $Y'Xa = X'Ya$; $a' = a$. Etant donné ces expressions, la fonction log-vraisemblance « $[3.11g]^c$ » se récrit finalement :

$$\begin{aligned}\ln L(a,\sigma_u^2) &= -\frac{T}{2}\ln\pi - \frac{T}{2}\ln(2\sigma_u^2) - \frac{1}{2\sigma_u^2}(Y-Xa)'(Y-Xa)\\ &= -\frac{T}{2}\ln\pi - \frac{T}{2}\ln(2\sigma_u^2) - \frac{1}{2\sigma_u^2}(Y'Y - 2X'aY + X'Xa^2) \dots\dots [3.11g]^d\end{aligned}$$

L'expression « $[3.11g]^d$ » est celle retenue pour l'estimation des paramètres.

Estimation des paramètres « a et σ_u^2 »

Pour estimer[4] les paramètres inconnus *« a et σ_u^2 » par la méthode du maximum de vraisemblance*, la démarche va consister à maximiser la fonction log-vraisemblance ci-dessus $[3.11g]^d$, ce qui revient à annuler ses dérivées premières par rapport aux arguments « a et σ_u^2 » comme suit :

En effet : $\left(\left[G[Y(x)]\right]^n\right)' = n\,G^{n-1}\,Y(x)'$; $\frac{d}{dX}[X^n] = n\,X^{n-1}$

$$\frac{\partial \ln L}{\partial a} = -\frac{1}{2\sigma_u^2}(-2X'Y + 2X'Xa) = 0 \dots\dots (i)$$

$$\frac{\partial \ln L}{\partial \sigma_u^2} = \frac{\partial}{\partial \sigma_u^2}\left(-\frac{T}{2}\ln(2\sigma_u^2)\right) + \frac{\partial}{\partial \sigma_u^2}\left[-\frac{1}{2\sigma_u^2}(Y'Y - 2X'aY + X'Xa^2)\right] = 0 \dots\dots (ii)$$

$$\rightarrow \frac{\boldsymbol{\partial}}{\boldsymbol{\partial}\boldsymbol{\sigma}_{\boldsymbol{u}}^{\mathbf{2}}}\left(-\frac{\boldsymbol{T}}{\mathbf{2}}\mathbf{ln}(\mathbf{2}\boldsymbol{\sigma}_{\boldsymbol{u}}^{\mathbf{2}})\right) = -\frac{T}{2\sigma_u^2}$$

$$\begin{aligned}\rightarrow \frac{\boldsymbol{\partial}}{\boldsymbol{\partial}\boldsymbol{\sigma}_{\boldsymbol{u}}^{\mathbf{2}}}\left[-\frac{\mathbf{1}}{\mathbf{2}}(\boldsymbol{\sigma}_{\boldsymbol{u}}^{\mathbf{2}})^{\mathbf{-1}}(\boldsymbol{Y}-\boldsymbol{Xa})'(\boldsymbol{Y}-\boldsymbol{Xa})\right] &= -\frac{1}{2} * (-1) * (\sigma_u^2)^{-1-1}(Y-Xa)'(Y-Xa)\\ &= \frac{1}{2\sigma_u^4}(Y-Xa)'(Y-Xa) = \frac{e'e}{2\sigma_u^4}\end{aligned}$$

Considérant ces expressions, (ii) s'écrit :

$$\frac{\partial \ln L}{\partial \sigma_u^2} = -\frac{T}{2\sigma_u^2} + \frac{1}{2\sigma_u^4}(Y-Xa)'(Y-Xa) = -\frac{T}{2\sigma_u^2} + \frac{e'e}{2\sigma_u^4} = \frac{-T\sigma_u^2 + e'e}{2\sigma_u^4} = 0 \dots\dots (ii)$$

Les expressions (i) et (ii) mises ensemble forment un système de « $k+1$ » équations dont la solution donne l'estimateur du maximum de vraisemblance des coefficients « $\hat{a}_{MV}$ » et celui de la variance de l'erreur « $\hat{\sigma}_{uMV}^2$ », soit respectivement :

[4] « La méthode du maximum de vraisemblance consiste à estimer les paramètres inconnus de manière que la probabilité d'observer les Y soit aussi forte que possible, c'est-à-dire soit maximale » (Bosonga B.L., p. 133).

$$\hat{a}_{MV} = (X'X)^{-1}X'Y \dots\dots [3.11g]^e$$

$$\hat{\sigma}^2_{uMV} = \frac{e'e}{T} \dots\dots [3.11g]^f$$

Au regard des résultats obtenus, soient les expressions $[3.11g]^e$ et $[3.11g]^f$, l'on note ce qui suit :

- L'estimateur MCO des paramètres « a » et l'estimateur du maximum de vraisemblance sont égaux : $\hat{a}_{MV} = \hat{a}_{MCO}$.
- Par contre, l'estimateur du maximum de vraisemblance « $\hat{\sigma}^2_{uMV}$ » de la variance de l'erreur « σ^2_u », étant biaisé, est différent de l'estimateur MCO, car (avec : $k = K - 1$ ou $K = k + 1$) :

$$E(\hat{\boldsymbol{\sigma}}^2_{\boldsymbol{u}\,\boldsymbol{MV}}) = \frac{1}{T}E\left(\sum_{t=1}^{T} e_t^2\right) \neq E(\hat{\boldsymbol{\sigma}}^2_{\boldsymbol{u}\,\boldsymbol{MCO}}) = \frac{\mathbf{1}}{(\boldsymbol{T}-\boldsymbol{k}-\mathbf{1})}E\left(\sum_{t=1}^{T} e_t^2\right) = \frac{\mathbf{1}}{(\boldsymbol{T}-\boldsymbol{K})} * (T-K)\sigma^2_u = \sigma^2_u$$

En effet :

$$E\left(\sum_{t=1}^{T} e_t^2\right) = (T-K)\sigma^2_u \rightarrow \boldsymbol{E}(\hat{\boldsymbol{\sigma}}^2_{\boldsymbol{u}\,\boldsymbol{MV}}) = \frac{1}{T} * (T-K)\sigma^2_u = \left(\frac{T-K}{T}\right)\sigma^2_u = \boldsymbol{\sigma}^2_{\boldsymbol{u}} - \frac{\boldsymbol{K}}{\boldsymbol{T}}\boldsymbol{\sigma}^2_{\boldsymbol{u}}$$

On constate que, l'estimateur MCO de la variance des erreurs étant sans biais, l'estimateur du maximum de vraisemblance de la variance des erreurs est biaisé vers le bas (en moyenne, il sous-estime la valeur réelle de σ^2_u), mais il reste convergent. Cette dernière propriété garantit la minimisation du biais avec l'accroissement de la taille de l'échantillon ; c'est-à-dire que, asymptotiquement (si T croit indéfiniment), $\hat{\sigma}^2_{u\,MV}$ est aussi non biaisé :

$$\lim_{\boldsymbol{T}\to\infty} \boldsymbol{E}(\hat{\boldsymbol{\sigma}}^2_{\boldsymbol{u}\,\boldsymbol{MV}}) = \boldsymbol{\sigma}^2_{\boldsymbol{u}} \;\; \boldsymbol{ou} \uparrow \boldsymbol{T} \rightarrow \hat{\boldsymbol{\sigma}}^2_{\boldsymbol{u}\,\boldsymbol{MV}} \simeq \hat{\boldsymbol{\sigma}}^2_{\boldsymbol{u}\,\boldsymbol{MCO}}$$

2. Quelques tests statistiques liés à l'estimateur du maximum de vraisemblance

Comme évoqué précédemment, nous présentons trois tests statistiques basés sur l'estimateur du maximum de vraisemblance, à savoir : le test de Wald (W), celui du ratio de vraisemblance ou Likelihood Ratio (LR) et celui du multiplicateur de Lagrange ou Lagrange Multiplier (LM).

Au sujet de ces tests, précisons ceci :

- Sur des grands échantillons ou lorsque la taille de l'échantillon s'accroit indéfiniment (asymptotiquement), contrairement aux petits échantillons, ces trois tests sont identiques ou mieux équivalents et leurs distributions statistiques suivent chacune la loi du khi-deux.
- Sous certaines conditions (hypothèses ou contraintes sur les coefficients d'un modèle linéaire), la relation suivante se vérifie : $W \geq LR \geq LM$; ceci signifie que l'hypothèse invalidée par le $LM\ test$ ne sera pas non plus acceptée ou validée par LR, ni par W.
- Dans ce cadre, les hypothèses qui font l'objet des tests sont les suivantes (*l'on test « p » contraintes linéaires sur un modèle qui est supposé être estimé par la méthode du maximum de vraisemblance*) :

$$H_0 = Ra - q = 0 \;\; ou \;\; Ra = q$$
$$H_1 = Ra - q \neq 0 \;\; ou \;\; Ra \neq q$$

Avec : $R =$ une matrice d'ordre « $p \times K$ » ; $p =$ vecteur colonne de format « $p \times 1$ » ; $p \leq K$; $a =$ coefficients du modèle ; $q =$ une valeur donnée.

Pour accepter ou rejeter l'hypothèse nulle, les règles de décision sont :

- Si la statistique calculée du test (Sc) est ≤ à sa valeur donnée sur la table (St), soit si la probabilité associée à la statistique calculée du test (Prob-test) est $\geq 5\%$: accepter H_0.
- Au cas contraire, c'est dire si $Sc > St \;\; ou \;\; Prob.test < 5\%$, rejeter H_0.

a) Le test de Wald (W)

- <u>La statistique du test</u> : Sous l'hypothèse nulle ($Ra = q$), la statistique « W » du test de Wald se présente comme suit ($\hat{a}_{MV}$ = *estimateur du maximum de vraisemblance des paramètres ou coefficients « a » ;* $\hat{\sigma}^2_{u\,MV}$ = *estimateur du maximum de vraisemblance de la variance des erreurs*) :

$$W = (R\hat{a}_{MV} - q)'[\hat{\sigma}^2_{u\,MV} R(X'X)^{-1}R'](R\hat{a}_{MV} - q) = p * F_c \sim \chi^2_{\alpha;p} \dots \dots [3.11h]$$

Avec : $p =$ nombre de contraintes sur les paramètres ou coefficients ; $F_c =$ statistique F de Fisher calculée ; $\chi^2_{\alpha;p} =$ distribution de khi-deux, au seuil de signification « α » de 5%, avec « p » degré de liberté. L'on note ainsi que la statistique de Wald est distribuée suivant un khi-deux à p degré de liberté : $W = p * F_c \sim \chi^2_{\alpha;p}$.

Sous H_0, à travers la statistique W ainsi construite, le test de Wald cherche à vérifier si l'écart entre « $R\hat{a}_{MV}$ » et « q » est significativement (statistiquement) égal à zéro.

- <u>Les règles de décision</u> : considérant la statistique W et les hypothèses du test, les règles de décision sont :
  - Accepter l'hypothèse nulle ($H_0 : Ra = q$), soit valider la contrainte, si : $F_c \leq \chi^2_{\alpha;p}$ ou si $Probabilité\ associée$ à $F_c \geq 5\%$.
  - Rejeter l'hypothèse nulle ($H_0 : Ra = q$), soit invalider la contrainte, si : $F_c \geq \chi^2_{\alpha;p}$ ou si $Probabilité\ associée$ à $F_c < 5\%$.

b) Le test du ratio de vraisemblance (LR test)

- <u>La statistique du test</u> : Sous l'hypothèse nulle ($Ra = q$), la statistique « LR » du test du ratio de vraisemblance est ($\hat{a}_{MV}$ = *estimateur du maximum de vraisemblance/ML des paramètres ou coefficients « a »*) :

$$LR = -2[\ln L(\hat{a}^*_{MV}) - \ln L(\hat{a}_{MV})] \sim \chi^2_{\alpha;p} \dots \dots [3.11h]^*$$

où :

$$\rightarrow \ln L(\hat{a}^*_{MV}) = -\frac{T}{2}\left[\ln(2\pi) + \ln\left(\hat{\boldsymbol{\sigma}}^{*2}_{\boldsymbol{u\,MV}}\right)\right] - \frac{e^{*\prime}e^*}{2\hat{\boldsymbol{\sigma}}^{*2}_{\boldsymbol{u\,MV}}}$$

$$\rightarrow \ln L(\hat{a}_{MV}) = -\frac{T}{2}\left[\ln(2\pi) + \ln\left(\hat{\boldsymbol{\sigma}}^{2}_{\boldsymbol{u\,MV}}\right)\right] - \frac{e'e}{2\hat{\boldsymbol{\sigma}}^{2}_{\boldsymbol{u\,MV}}}$$

L'expression $[3.11h]^*$ peut aussi s'écrire :

$$LR = T[\ln L(e^{*\prime}e^*) - \ln L(e'e)] \sim \chi^2_{\alpha;p} \dots \dots [3.11h]^{**}$$

Avec : $\hat{a}^*_{MV}$ = estimateur ML des paramètres ou coefficients du « modèle contraint » ; $\hat{a}_{MV}$ = estimateur ML des paramètres ou coefficients du « modèle non contraint » ; $e^{*\prime}e^*$ et $e'e$ = respectivement, somme des carrés des résidus du modèle contraint et celle du modèle non contraint ; $\chi^2_{\alpha;p}$ = distribution de khi-deux, au seuil de signification « α » de 5%, avec « p » degré de liberté.

L'on note que la statistique du ratio de vraisemblance est distribuée suivant un khi-deux à p degré de liberté : $LR \sim \chi^2_{\alpha;p}$.

Sous H_0, à travers la statistique LR ainsi construite, le test du ratio de vraisemblance compare les valeurs du logarithme de vraisemblance (Log Likelihood) de deux modèles estimés : le modèle contraint (soit le modèle sous H_0) et le modèle non contraint.

- Les règles de décision : considérant la statistique LR et les hypothèses du test, les règles de décision sont :
  - Accepter l'hypothèse nulle ($H_0 : Ra = q$), soit valider la contrainte, si : $LR \leq \chi^2_{\alpha;p}$ ou si $Probabilit\acute{e}\ associ\acute{e}e$ à $LR \geq 5\%$.
  - Rejeter l'hypothèse nulle ($H_0 : Ra = q$), soit invalider la contrainte, si : $LR \geq \chi^2_{\alpha;p}$ ou si $Probabilit\acute{e}\ associ\acute{e}e$ à $LR < 5\%$.

c) Le test du multiplicateur de Lagrange (LM test) ou test du Score

- La statistique du test : Sous l'hypothèse nulle ($Ra = q$), la statistique « LM » du test du multiplicateur de Lagrange est ($\hat{a}_{MV}$ = *estimateur du maximum de vraisemblance/ML des paramètres ou coefficients « a »*) :

$$LM = \frac{T}{e^{*\prime}e^*}(X'e^*)'(X'X)^{-1}(X'e^*) = T * R^2_e \sim \chi^2_{\alpha;p} \dots \dots [3.11i]$$

Avec : $e^{*\prime}e^*$ et $e'e$ = respectivement, somme des carrés des résidus du modèle contraint (modèle sous H_0) et celle du modèle non contraint ; $\chi^2_{\alpha;p}$ = distribution de khi-deux, au seuil de signification « α » de 5%, avec « p » degré de liberté ; T = taille de l'échantillon ; R^2_e = coefficient de détermination de l'estimation par les MCO du modèle ou de la relation suivante : $e^* = f(X)$, avec : X = les variables explicatives.

L'on note que la statistique du multiplicateur de Lagrange, basée sur les informations de la condition du premier ordre de maximisation de la fonction log vraisemblance, est distribuée suivant un khi-deux à p degré de liberté : $LM = TR^2_e \sim \chi^2_{\alpha;p}$.

Sous H_0, à travers la statistique LM ainsi construite, le test du multiplicateur de Lagrange cherche à vérifier que les multiplicateurs de Lagrange associés aux restrictions ou contraintes définies (Cfr H_0) approchent zéro ; ce qui signifie, dans ce cas, que la perte de vraisemblance est faible (ou que la vraisemblance est maximale) pour les contraintes imposées sous H_0.

- Les règles de décision : considérant la statistique LM et les hypothèses du test, les règles de décision sont :
  - Accepter l'hypothèse nulle ($H_0 : Ra = q$), soit valider la contrainte, si : $LM \leq \chi^2_{\alpha;p}$ ou si $Probabilit\acute{e}\ associ\acute{e}e$ à $LM \geq 5\%$.
  - Rejeter l'hypothèse nulle ($H_0 : Ra = q$), soit invalider la contrainte, si : $LM \geq \chi^2_{\alpha;p}$ ou si $Probabilit\acute{e}\ associ\acute{e}e$ à $LM < 5\%$.

IIème Partie
Pratiques sur Logiciel (Eviews 10)

Méthode du maximum de vraisemblance (MV) : estimation et tests d'hypothèses

a) Estimation par le MV

Exemple 1 : Estimation de la probabilité de recourir au centre de santé en cas de paludisme

Soit la relation suivante :

$$Y_i = a_0 + a_1 X_i + u_i \dots \dots [3.11j]$$

Avec :

- $Y_i =$ mode de traitement, par le ménage i, en cas de survenance du paludisme. Cette variable dépendante est binaire, avec les modalités : 1 = recourir à un centre médical ; 0 = recourir à l'automédication.
- X_i : revenu hebdomadaire du ménage i (en usd).
- a_0 et a_1 : les paramètres du modèle.

Sur base de données disponibles, observées sur 40 ménages, nous allons estimer la relation ci-dessus (soit le modèle Logit) par la méthode du maximum de vraisemblance (MV).

Pour ce faire, sur Eviews, les commandes sont :

create u 1 40
data Yi Xi
copier et coller les données (Ctrl+C et Ctrl+V)
suivre : Quick/Estimate Equation…→ dans la boite de dialogue, taper : **Yi c Xi** → dans Method, sélectionner : BINARY – Binary Choice (Logit, Probit, Extrem Value) → cocher sur **Logit** → ensuite cliquer sur **OK**.

Les résultats sont :

Dependent Variable: YI
Method: ML - Binary Logit (Newton-Raphson / Marquardt steps)
Date: 06/07/19 Time: 17:28
Sample: 1 40
Included observations: 40
Convergence achieved after 5 iterations
Coefficient covariance computed using observed Hessian

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	3.532044	1.770798	1.994606	0.0461
XI	-0.240169	0.112755	-2.130013	0.0332

McFadden R-squared	0.127433	Mean dependent var	0.425000
S.D. dependent var	0.500641	S.E. of regression	0.470578
Akaike info criterion	1.289927	Sum squared resid	8.414876
Schwarz criterion	1.374371	Log likelihood	-23.79854
Hannan-Quinn criter.	1.320459	Deviance	47.59709
Restr. deviance	54.54837	Restr. log likelihood	-27.27418
LR statistic	6.951283	Avg. log likelihood	-0.594964
Prob(LR statistic)	0.008376		

Obs with Dep=0	23	Total obs	40
Obs with Dep=1	17		

Estimation Command :
=========================
BINARY(D=L) YI C XI

Estimation Equation :
=========================
I_YI = C(1) + C(2)*XI

Forecasting Equation :
=========================
YI = 1-@CLOGISTIC(-(C(1) + C(2)*XI))

Substituted Coefficients :
=========================
YI = 1-@CLOGISTIC(-(3.53204397397 - 0.240168826364*XI))

b) ***Tests d'hypothèse basés sur l'estimateur du maximum de vraisemblance***

➢ ***Test de WALD***

Testons la contrainte (restriction) suivante : $a_0 = -a_1$, ce qui revient à tester les hypothèses suivantes :

$H_0 = a_0 + a_1 = 0$
$H_1 = a_0 + a_1 \neq 0$

Pour ce faire, sur Eviews, dans l'output de l'estimation, suivre : *View/Coefficient Diagnostics/Wald-Coefficient Restrictions…→ taper : c(1)+c(2)=0→ ok* **:**

Wald Test:			
Equation: Untitled			
Test Statistic	Value	df	Probability
t-statistic	1.982599	38	0.0547
F-statistic	3.930699	(1, 38)	0.0547
Chi-square	**3.930699**	**1**	**0.0474**
Null Hypothesis: C(1)+C(2)=0			
Null Hypothesis Summary:			
Normalized Restriction (= 0)		Value	Std. Err.
C(1) + C(2)		3.291875	1.660384
Restrictions are linear in coefficients.			

On sait que la statistique de Wald suit une distribution limite de khi-deux à p degré de liberté, au seuil de signification « α » de 5% : $W = p * F_c \sim \chi^2_{\alpha;p}$. Aussi, suivant les règles de décision, l'on rejette l'hypothèse nulle ou invalide la contrainte si : $F_c \geq \chi^2_{\alpha;p}$ ou si $Probabilité\ associée\ à\ la\ statistique\ calculée < 5\%$. Dans notre cas, cette probabilité (0,0474) est $< 5\%$, d'où : $a_0 + a_1 \neq 0$.

- ***Test du ratio de vraisemblance***

Testons la contrainte (restriction) suivante : $a_0 = -\, a_1$, ce qui revient à tester les hypothèses suivantes :

$H_0 = a_0 + \ a_1 = 0$: non significativité conjointe des paramètres estimés.
$H_1 = a_0 + \ a_1 \neq 0$: significativité conjointe des paramètres estimés.

Pour ce faire, référons nous aux résultats d'estimation qui nous fournissent les informations suivantes **:**

LR statistic	6,951283
Prob(LR statistic)	0,008376

On sait que la statistique du ratio de vraisemblance suit une distribution limite de khi-deux à p degré de liberté, au seuil de signification « α » de 5% : $LR \sim \chi^2_{\alpha;p}$. Aussi, suivant les règles de décision, l'on rejette l'hypothèse nulle ou invalide la contrainte si : $LR \geq \chi^2_{\alpha;p}$ ou si $Probabilit\acute{e}\ associ\acute{e}e$ à $LR < 5\%$. Dans notre cas, cette probabilité (0,008376) est $< 5\%$, d'où : $a_0 + \ a_1 \neq 0$.

- ***Test du multiplicateur de Lagrange***

Ce test n'est pas disponible dans Eviews pour des modèles probit/logit. Toutefois, LM test permettant aussi de tester l'autocorrélation des erreurs, l'on peut s'en servir pour tester l'hypothèse d'autocorrélation des erreurs dans le modèle $[3.11k]$ ci-dessous :

$$\ln Q_i = \ln A + b_1 \ln K_i + b_2 \ln L_i + u_i \rightarrow q_i = b_0 + b_1 k_i + b_2 l_i + u_i \ldots \ldots [3.11k]$$

$$e_i = a_0 + a_1 k_i + a_2 l_i + \boldsymbol{a_3 e_{i(-1)}} + \boldsymbol{a_4 e_{i(-2)}} \ldots \ldots [3.11l]$$

Avec : $e_i = q_i - \hat{q}_i$

Il sera question de tester la contrainte « $a_3 = 0\ et\ \ a_4 = 0$ » sur la relation $[3.11l]$, soit tester les hypothèses suivantes (MC = modèle contraint ; MNC = modèle non contraint) :

$H_0 :\ a_3 = 0\ et\ \ a_4 = 0 \rightarrow MC :\ e_i = a_0 + a_1 k_i + a_2 l_i$
$H_1 :\ a_3 \neq 0\ et\ a_4 \neq 0 \rightarrow MNC : e_i = a_0 + a_1 k_i + a_2 l_i + \boldsymbol{a_3 e_{i(-1)}} + \boldsymbol{a_4 e_{i(-2)}}$

Rejeter H_0, c'est admettre l'autocorrélation des erreurs.

Pour réaliser ce test sur Eviews, dans l'output de l'estimation, suivre : *View/Residual Diagnostics/Serial Correlation LM Test…→ Lags to include, taper : 2→ ok* **:**

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	9.886676	Prob. F(2,5)	0.0183
Obs*R-squared	**7.981702**	**Prob. Chi-Square(2)**	**0.0185**

Test Equation:
Dependent Variable: RESID
Method: Least Squares
Date: 06/07/19 Time: 18:21
Sample: 1 10
Included observations: 10
Presample missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.343923	0.159391	-2.157731	0.0834
LK	0.146508	0.074245	1.973306	0.1055
LL	-0.028433	0.039180	-0.725693	0.5006
RESID(-1)	**-0.595229**	**0.251280**	**-2.368786**	**0.0640**
RESID(-2)	**-1.175086**	**0.274806**	**-4.276055**	**0.0079**

R-squared	0.798170	Mean dependent var	6.84E-16
Adjusted R-squared	0.636706	S.D. dependent var	0.068425
S.E. of regression	0.041242	Akaike info criterion	-3.231843
Sum squared resid	0.008505	Schwarz criterion	-3.080550
Log likelihood	21.15921	Hannan-Quinn criter.	-3.397810
F-statistic	4.943338	Durbin-Watson stat	2.249080
Prob(F-statistic)	0.054818		

On sait que la statistique du multiplicateur de Lagrange, qui se base sur les informations de la condition du premier ordre de maximisation de la fonction log vraisemblance, suit une distribution limite de khi-deux à p degré de liberté, au seuil de signification « α » de 5% : $LM = TR_e^2 \sim \chi^2_{\alpha;p}$. Aussi, suivant les règles de décision, l'on rejette l'hypothèse nulle ou invalide la contrainte si : $LM \geq \chi^2_{\alpha;p}$ ou si $Probabilité\ associée$ à $LM < 5\%$.. Dans notre cas, cette probabilité (0,0185) est $< 5\%$, d'où : $a_3 \neq 0\ et\ a_4 \neq 0$ (il y a autocorrélation des erreurs).

Références bibliographiques

➢ Ouvrages et articles

Baraud Y. et Birgé L. (2017), « *Une alternative robuste au maximum de vraisemblance : la ρ-estimation* », in Journal de la Société Française de Statistique (http://www.sfds.asso.fr/journal), Vol. 158, No. 3., pp. 1-26.

Bertaux N. et al. (1999), « *Maximum de vraisemblance paramétrée (PML) : application à la localisation spatio-temporelle radar* », in Traitement du Signal, Volume 16 – n°3, pp. 187-201.

Bosonga Bofeki L. JP. (2019), « *Manuel d'Econométrie* », PUK, Editions Terabytes, RDC, 326 p.

Bourbonnais R. (2015), « *Econométrie : cours et exercices corrigés* », 9è édition, éd. DUNOD, Paris, 392 p.

______________(2009), « *Logiciel Eviews* », Université de Paris-Dauphine, 31 p.

Danglos Grégory (2009), « *Introduction à l'économétrie* », PUF, Paris, 245 p.

Foulley J-L., Delmas C. et Robert-Granié C. (2002), « *Méthodes du maximum de vraisemblance en modèle linéaire mixte* », in Journal de la Société Française de Statistique, tome 143, n°1-2, pp. 5-52.

Huber Catherine (1997), « *Remarques sur le maximum de vraisemblance* », in Questiio, vol. 21, 1 i 2, pp. 37-58.

Kibala Kuma J. (2019), « *L'Econométrie avec Eviews : Répertoire de quelques commandes de base* », Editions Universitaires Européennes, 117 p.

Kintambu Mafuku E.G. (2004), « *Principes d'Econométrie* », Presses de l'Université Kongo, 4è édition, 285 p.

Le cadre Jean-Pierre (1985), « *Estimation au sens du maximum de vraisemblance d'un modèle autorégressif. Application au traitement spatial d'antennes* », in Traitement du signal, Volume 2-n°4, pp. 291-303.

Mignon Valérie (2008), « *Econométrie : Théorie et Applications* », éd. ECONOMICA, 236 p.

➢ Séminaires, Rapports de recherche et Notes des cours

Bailly-Bechet M., « *Vraisemblance* », Notes de cours Biostatistiques – MIV (L3), Université Claude Bernard Lyon 1 – France, 15 p.

Berger Marie-Odile (2018), « *Méthodes d'estimation* », Notes du 13 Novembre (http://members.loria.fr/moberger).

Chesneau Christophe (2017), « *Sur l'Estimateur du Maximum de Vraisemblance* », Licence (France), 68 p. (publié dans Hal en ligne : https://cel.archives-ouvertes.fr/cel-01430435).

Deschamps Philippe (2006), « *Cours d'Econométrie* », Séminaire d'Econométrie, Université de Fribourg (Suisse), 258 p.

Detais Amélie (2008), « *Maximum de Vraisemblance et Moindres carrés pénalisés dans des modèles de durées de vie censurées* », Thèse, Université Toulouse III, 148 p.

Dufour Jean-Marie (2002), « *Estimation de modèles ARMA par la méthode du maximum de vraisemblance* », Université de Montréal, 9 p.

El Asri Mohammed, « *Etudes des M-estimateurs et leurs versions pondérées pour des données clustérisées* », Thèse défendue à l'Université d'Avignon et des pays de Vaucluse, 113 p.

Gassiat Elisabet, « *Statistiques* », Notes de cours – M1, 100 p.

Gaudoin Olivier, « *Statistique Inférentielle Avancée* », Notes de cours, 135 p.

Keribin Christine (2018), « *Cours 2 : Estimation par Maximum de Vraisemblance* », M1-Mathématiques Appliquées (Modélisation Statistique), Université Paris Sud.

Lemonnier Florian, « *Estimateur du maximum de vraisemblance pour le paramètre d'une loi* $u([0,\theta])$ », ENS Rennes – Université de Rennes 1.

Mariadassou M., Brault V. et Keribin C. (2017), « *Normalité asymptotique de l'estimateur du maximum de vraisemblance dans le modèle de blocs latents* », 48èmes Journées de Statistique de la SFdS, May 2016, Montpellier (France), 7 p. (publié dans Hal en ligne : https://hal.inria.fr/hal-01440084).

Salloum Zahraa (2018), « *Maximum de vraisemblance empirique pour la détection de changements dans un modèle avec un nombre faible ou très grand de variables* ». Université de Lyon, Thèse défendue en 2016, 183 p. (publié dans Hal en ligne : https://tel.archives-ouvertes.fr/tel-01360040).

Saumard Matthieu (2016), « *Maximum de Vraisemblance, Information de Fisher* ».

Tisserant S. (2009), « *Eléments de Statistiques* », 21 p.

Trevezas Samis (2010), « *Etude de l'estimation du Maximum de Vraisemblance dans des modèles Markoviens, Semi-Markoviens et Semi-Markoviens Cachés avec Applications* », Thèse défendue à l'Université de Technologie de Compiègne en 2008 (publiée dans Hal en ligne : https://tel.archives-ouvertes.fr/tel-00472644), 159 p.

Viano M-C. et Suquet C. (2010), « *Eléments de Statistique Asymptotique* », Master 2 Recherche de Mathématiques, 82 p.

Zaidi Abdelhamid (2004), « *Séparation aveugle d'un mélange instantané de sources autorégressives par la méthode du maximum de vraisemblance exact* », Laboratoire de Modélisation et Calcul – UMR 5523. Université Joseph-Fourier – Grenoble I, Thèse défendue en 2000, 133 p. (publié dans Hal en ligne : https://tel.archives-ouvertes.fr/tel-00006754).
