# Section 02 : Régression et Moindres Carrés Ordinaires (Séance 3)

## GSF-6053: Économétrie Financière

Simon-Pierre Boucher[1]

[1]Département de finance, assurance et immobilier
Faculté des sciences de l'administration
Université Laval

25 Janvier 2022

# Références

**Obligatoires:**

- **Notes de cours:** Section 2 (Professeure: Marie-Hélène Gagnon)
- **Woolridge:** chapitres 2 à 7

**Complémentaires:**

- **Gujarati et Porter:** chapitres 1 à 9.
- **Greene:** chapitres 2, 3, 4, 5, 9, 14, 20, appendices C et D

# Plan de la séance

Maximum de Vraissemblance

MLE: Modèle de régression linéaire

Propriétés Estimateur MLE ET MCO

Propriétés de $\hat{\sigma}^2$

Estimateur de la variance sans biais

# Maximum de Vraissemblance

- Le maximum de vraisemblance est une méthode générale pour estimer les paramàtres d'un modèle statistique.
- Nous aurons une serie d'observations d'une variable aléatoire y et un modèle statistique potentiel pour cette variable.
  - Ce modèle peut inclure la dépendance de y sur d'autres variables prédictrices.
  - Ainsi qu'une distribution statistique pour la portion non-expliquée de la variation de y.
- Selon le maximum de vraisemblance, les meilleurs estimés des paramètres d'un modèle sont ceux qui maximisent la probabilité des valeurs observées de la variable

# Maximum de Vraissemblance

- Fonction de densité pour une variable aléatoire $Y$ conditionné sur un ensemble de paramètres $\theta$

$$f(Y \mid \theta)$$

- La fonction de densité jointe de $n$ observations est simplement le produit des densités individuels

$$f(y_1, y_2, \ldots, y_n \mid \theta) = \prod_{i=1}^{n} f(y_i \mid \theta) = L(\theta \mid Y)$$

  - Ensemble de paramètres $\theta$ est inconnu.
  - Nous voulons les valeurs de $Y$ provenant de l'échantillon.
- Donner à $\theta$ la valeur qui maximise la probabilité d'obtenir un échantillon identique à celui qu'on dispose.

# MLE: Modèle de régression linéaire

- Hypothèse distributionnelle sur les $u_t$

$$u_t \sim \; iid \; N(0, \sigma^2)$$

- Loi conjointe des $u_1, u_2, ..., u_T$

$$f(u_1, u_2, ..., u_T) = \prod_{t=1}^{T} f(u_t) = \left( \frac{1}{\sqrt{2\pi\sigma^2}} \right)^T e^{-\frac{\sum_{t=1}^{T}(u_t)^2}{2\sigma^2}}$$

- Interessés à la loi coinjointes des $y_t$
- Effectué un changement de variable:

$$u_t = Y_t - X_t'\beta$$

# MLE: Modèle de régression linéaire

- Loi conjointe des $Y_t$ (Vraissemblance)

$$f(Y_1, Y_2, ..., Y_T) = \prod_{t=1}^{T} g(Y_t) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^T e^{-\frac{\sum_{t=1}^{T}(Y_t - X_t'\beta)^2}{2\sigma^2}}$$

$$= (2\pi\sigma^2)^{-T/2} e^{-\frac{(Y-X\beta)'(Y-X\beta)}{2\sigma^2}}$$

- On veut enseuite obtenir la Log-Vraissemblance en prenant le logarithme de la fonction de vraissemblance.

$$L = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log(\sigma^2) - \frac{1}{2} \times \frac{(Y - X\beta)'(Y - X\beta)}{\sigma^2}$$

# MLE: Modèle de régression linéaire

- Maximiser la log-vraissemblance en fonction de $\beta$ et $\sigma^2$
- **En fonction de** $\beta$

$$\frac{\partial L}{\partial \beta} = \frac{1}{\sigma^2} \times \frac{\partial(Y'Y = 2Y'X\beta + \beta'X'X\beta)}{\partial \beta} = 0$$

**Condition de première ordre:**

$$-2X'Y + 2X'X\beta = 0$$

$$-X'Y + X'X\beta = 0$$

$$X'X\beta = X'Y$$

**Estimateur** $\beta$ **par MLE:**

$$\hat{\beta} = (X'X)^{-1}X'Y$$

# MLE: Modèle de régression linéaire

- **En fonction de $\sigma^2$**

$$\frac{\partial L}{\partial \sigma^2} = \frac{\partial \left[ -\frac{T}{2} \log(\sigma^2) - \frac{1}{2} \times \frac{(Y - X\beta)'(Y - X\beta)}{\sigma^2} \right]}{\partial \sigma^2} = 0$$

- Nous allons dériver par rapport à $\sigma^2$ en deux parties
  - 1er terme:

$$\frac{\partial \left[ -\frac{T}{2} \log(\sigma^2) \right]}{\partial \sigma^2} = -\frac{T}{2} \times \frac{1}{\sigma^2}$$

  - 2e terme:

$$\frac{\partial \left[ -\frac{1}{2} \times \frac{(Y - X\beta)'(Y - X\beta)}{\sigma^2} \right]}{\partial \sigma^2} = \frac{\partial \left[ -\frac{1}{2} \times (\sigma^2)^{-1}(Y - X\beta)'(Y - X\beta) \right]}{\partial \sigma^2}$$

# MLE: Modèle de régression linéaire

- **En fonction de $\sigma^2$**
  - Suite pour le 2e terme:

$$\begin{aligned}
&\frac{\partial\left[-\frac{1}{2}\times(\sigma^2)^{-1}(Y-X\beta)'(Y-X\beta)\right]}{\partial\sigma^2}\\
&=\frac{T}{2}\times(\sigma^2)^{-2}(Y-X\beta)'(Y-X\beta)\\
&=\frac{1}{2}\times\sigma^{-4}(Y-X\beta)'(Y-X\beta)\\
&=\frac{T}{2}\times\frac{(Y-X\beta)'(Y-X\beta)}{\sigma^4}
\end{aligned}$$

# MLE: Modèle de régression linéaire

- **En fonction de $\sigma^2$**
  - On additionne les deux termes et nous avons notre condition de première ordre

$$\frac{\partial L}{\partial \sigma^2} = -\frac{T}{2} \times \frac{1}{\sigma^2} + \frac{1}{2} \times \frac{(Y - X\beta)'(Y - X\beta)}{\sigma^4}$$

- On veut maintenant isoler $\sigma^2$ pour obtenir l'estimateur $\hat{\sigma}^2$

$$\frac{1}{2} \times \frac{(Y - X\beta)'(Y - X\beta)}{\sigma^4} = \frac{T}{2} \times \frac{1}{\sigma^2}$$

- On peut multiplier par $2\sigma^2$ de chaque coté pour simplifier

$$\frac{(Y - X\beta)'(Y - X\beta)}{\sigma^2} = T$$

**Estimateur $\hat{\sigma}^2$ pour la méthode des MLE**

$$\hat{\sigma}^2 = \frac{(Y - X\beta)'(Y - X\beta)}{T}$$

Maximum de Vraissemblance

MLE: Modèle de régression linéaire

Propriétés Estimateur MLE ET MCO

Propriétés de $\hat{\sigma}^2$

Estimateur de la variance sans biais

# Propriétés Estimateur MLE ET MCO

- Pour $\hat{\beta}$ des MCO et MLE, leurs résultats coincident

## Estimateur **BLUE**

- Best linear unbiased estimator
  - Estimateur sans biais
  - Estimateur ayant une variance minimal
- L'estimateur des moindres carrés ordinaires est BLUE
- L'estimateur $\hat{\sigma}^2$ du Maximum de vraissemblance est biaisé vers le bas
- On peut trouver une alternative sans biais

$$\hat{S}^2 = \frac{(Y - X\hat{\beta})'(Y - X\hat{\beta})}{T - K}$$

# Propriétés Estimateur MLE ET MCO

## Estimateur **BLUE**

- **Sans biais** : Espérance de l'estimateur égale à la vraie valeur du paramètre

$$E(\hat{\theta}) = \theta$$

- **Efficace:** Si l'estimateur atteint la borne de Crameur-Rao (autrement dit, l'inverse de la matrice d'information de fisher)
- On veut un estimateur ayant la variance la plus petite possible
  - Cela donne une meilleur pécision

## Propriétés Estimateur MLE ET MCO

- Borne de Cramer-Rao : Pour tout estimateur regulier et sans biais, sa variance est bornee par l'inverse de la matrice d'information.
- La matrice d'information est quant a elle une facon de mesurer la quantite d'information sur les parametres dans $\theta$ contenue dans $X$.
- Une definition equivalente serait que la variance d'un estimateur sans biais sera toujours au moins aussi grande que l'inverse de la matrice d'information :

### Inverse de la matrice d'information

$$
\begin{aligned}
[I(\theta)]^{-1} &= \left(-E\left[\frac{\partial^2 \log L(\theta)}{\partial \theta^2}\right]\right)^{-1} \\
&= \left(E\left[\left(\frac{\partial \log L(\theta)}{\partial \theta}\right)^2\right]\right)^{-1}
\end{aligned}
$$

# Propriétés Estimateur MLE ET MCO

## Matrice d'information et Hessienne

- La matrice d'information est simplement une matrice hessienne d'une d'une fonction.
- Il s'agit essentiellement d'une matrice de dérivé seconde:
- On suppose une fonction $f(x_1, x_2)$ et on représente la hessienne de cette fonction par $H_{i,j}(f)$
- On doit représenter $H_{i,j}(f)$ comme étant l'ensemble des dérivés secondes partiels possible.

$$H_{i,j} = \frac{\partial^2 f}{\partial x_i \partial x_j}$$

- Il y aura donc 4 dérivés secondes partiels possibles
  - $i = 1$ et $j = 1$, alors $\partial x_1^2$
  - $i = 1$ et $j = 2$, alors $\partial x_1 \partial x_2$
  - $i = 2$ et $j = 1$, alors $\partial x_2 \partial x_1$
  - $i = 2$ et $j = 2$, alors $\partial x_2^2$

# Propriétés Estimateur MLE ET MCO

## Matrice d'information

**Dans le cas du modèle linéaire estimé par MLE**

- On aura 4 dérivés seconde étant donnée que nous avons deux paramètres à estimer, soit $\beta$ et $\sigma^2$.
  - **En haut à gauche:** $\partial\beta\partial\beta$
  - **En haut à droite:** $\partial\beta\partial\sigma^2$
  - **En bas à gauche:** $\partial\sigma^2\partial\beta$
  - **En bas à droite:** $(\partial\sigma^2)^2$

$$I(\beta, \sigma^2) = -E\begin{bmatrix} \left(\frac{\partial^2 L}{\partial\beta\partial\beta'}\right) & \left(\frac{\partial^2 L}{\partial\beta\partial\sigma^2}\right) \\ \left(\frac{\partial^2 L}{\partial\sigma^2\partial\beta'}\right) & \left(\frac{\partial^2 L}{(\partial\sigma^2)^2}\right) \end{bmatrix}$$

- Nous allons maintenant résoudre les 4 dérivés secondes partiels possibles:

# Propriétés Estimateur MLE ET MCO

## En haut à gauche

$$\frac{\partial^2 L}{\partial\beta\partial\beta'} = \frac{\partial^2\left[-\frac{T}{2}\log(2\pi) - \frac{T}{2}\log(\sigma^2) - \frac{1}{2}\times\frac{(Y-X\beta)'(Y-X\beta)}{\sigma^2}\right]}{\partial\beta\partial\beta'}$$

$$= \frac{\partial\left[-\frac{1}{2\sigma^2}\times(-2X'Y + 2X'X\hat{\beta})\right]}{\partial\beta}$$

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

# Propriétés Estimateur MLE ET MCO

## En haut à droite

$$\frac{\partial^2 L}{\partial\beta\partial\sigma^2} = \frac{\partial^2 \left[-\frac{T}{2}\log(2\pi) - \frac{T}{2}\log(\sigma^2) - \frac{1}{2} \times \frac{(Y-X\beta)'(Y-X\beta)}{\sigma^2}\right]}{\partial\beta\partial\sigma^2}$$

$$= \frac{\partial\left[-\frac{1}{2\sigma^2} \times (-2X'Y + 2X'X\hat{\beta})\right]}{\partial\sigma^2}$$

$$= \frac{1}{2\sigma^4} \times (-2X'Y + 2X'X\hat{\beta})$$

$$= \frac{1}{\sigma^4} \times (X'Y - X'X\hat{\beta})$$

$$= -\frac{1}{\sigma^4} \times (X'[Y - X\hat{\beta}])$$

$$= -\frac{1}{\sigma^4}(X'u)$$

Sachant $Y - X\hat{\beta} = u$

# Propriétés Estimateur MLE ET MCO

En bas à gauche

$$\frac{\partial^2 L}{\partial\sigma^2\partial\beta'} = \left(\frac{\partial^2 L}{\partial\beta\partial\sigma^2}\right)'$$
$$= -\frac{1}{\sigma^4}(X'u)'$$
$$= -\frac{1}{\sigma^4}(u'X)$$

# Propriétés Estimateur MLE ET MCO

## En bas à droite

$$
\begin{aligned}
\frac{\partial^2 L}{(\partial\sigma^2)^2} &= \frac{\partial^2\left[-\frac{T}{2}\log(2\pi) - \frac{T}{2}\log(\sigma^2) - \frac{1}{2}\times\frac{(Y-X\beta)'(Y-X\beta)}{\sigma^2}\right]}{(\partial\sigma^2)^2} \\
&= \frac{\partial\left[-\frac{-T}{2\sigma^2} + \frac{(Y-X\beta)'(Y-X\beta)}{2\sigma^4}\right]}{\partial\sigma^2} \\
&= \frac{T}{2\sigma^4} - \frac{(Y-X\beta)'(Y-X\beta)}{\sigma^6}
\end{aligned}
$$

# Propriétés Estimateur MLE ET MCO

- Espérance mathématique de chacune des dérivés

**En haut à gauche**

$$E\left(-\frac{\partial^2 L}{\partial\beta\partial\beta'}\right) = E\left(-\left[-\frac{1}{\sigma^2}(X'X)\right]\right) = \frac{1}{\sigma^2}(X'X)$$

# Propriétés Estimateur MLE ET MCO

## En haut à droite

$$
\begin{aligned}
E\left(-\frac{\partial^2 L}{\partial\beta\partial\sigma^2}\right) &= E\left(-\left[-\frac{1}{\sigma^4}(X'Y - X'X\beta)\right]\right) \\
&= \frac{1}{\sigma^4}(X'E(Y) - X'X\beta) \\
&= \frac{1}{\sigma^4}(X'X\beta - X'X\beta) \\
&= 0
\end{aligned}
$$

Sachant $E(Y) = X\beta$

# Propriétés Estimateur MLE ET MCO

En bas à gauche

$$E\left(-\frac{\partial^2 L}{\partial^2 \partial\beta'}\right) = 0$$

En bas à droite

$$\begin{aligned}E\left(-\frac{\partial^2 L}{(\partial^2)^2}\right) &= E\left(-\left[\frac{T}{2\sigma^4} - \frac{(Y-X\beta)'(Y-X\beta)}{\sigma^6}\right]\right)\\ &= -\frac{T}{2\sigma^4} + \frac{E[(Y-X\beta)'(Y-X\beta)]}{\sigma^6}\end{aligned}$$

# Propriétés Estimateur MLE ET MCO

En bas à droite

**Utilisons la trace:**

$$
\begin{aligned}
E[(Y - X\beta)'(Y - X\beta)] &= E(u'u) \\
&= E(Trace(u'u)) \\
&= E(Trace(uu')) \\
&= Trace(E(uu')) \\
&= Trace(\sigma^2 I_T) \\
&= T\sigma^2
\end{aligned}
$$

# Propriétés Estimateur MLE ET MCO

## En bas à droite

**Donc:**

$$\begin{aligned}
E\left(-\frac{\partial^2 L}{(\partial^2)^2}\right) &= -\frac{T}{2\sigma^4} + \frac{T\sigma^2}{\sigma^6} \\
&= -\frac{T}{2\sigma^4} + \frac{T}{\sigma^4} \\
&= -\frac{T}{2\sigma^4} + \frac{2T}{2\sigma^4} \\
&= \frac{T}{2\sigma^4}
\end{aligned}$$

# Propriétés Estimateur MLE ET MCO

Matrice d'information

$$I(\beta, \sigma^2) = \begin{bmatrix} \frac{1}{\sigma^2}(X'X) & 0 \\ 0 & \frac{T}{2\sigma^4} \end{bmatrix}$$

Inverse matrice d'information

$$I^{-1}(\beta, \sigma^2) = \begin{bmatrix} \sigma^2(X'X) & 0 \\ 0 & \frac{2\sigma^4}{T} \end{bmatrix}$$

# Propriétés Estimateur MLE ET MCO

## Espérance

On sait déja que l'estimateur $\hat{\beta}$ possède la solution suivante:

$$\hat{\beta} = (X'X)^{-1}X'Y$$

Sachant $Y = X\beta + u$

$$\hat{\beta} = (X'X)^{-1}X'[X\beta + u]$$

$$\hat{\beta} = (X'X)^{-1}X'X\beta + (X'X)^{-1}X'u$$

Sachant également $(X'X)^{-1}X'X = I$

$$\hat{\beta} = \beta + (X'X)^{-1}X'u$$

# Propriétés Estimateur MLE ET MCO

## Espérance

On applique l'espérance de chaque coté de l'équation

$$\begin{aligned} E(\hat{\beta}) &= E(\beta + (X'X)^{-1}X'u) \\ &= \beta + (X'X)^{-1}X'E(u) \end{aligned}$$

Sachant $E(u) = 0$

$$E(\hat{\beta}) = \beta$$

**On peut donc maintenant affirmer que $\hat{\beta}$ est un estimateur sans biais de $\beta$**

# Propriétés Estimateur MLE ET MCO

## Variance

On peut exprimer la variance de $\hat{\beta}$ comme suit:

$$Var(\hat{\beta}) = E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)']$$

Sachant l'éqation que nous avons déja obtenus dans le calcule de l'espérance:

$$\hat{\beta} = \beta + (X'X)^{-1}X'u$$

Alors il nous est possible d'exprimer la déviation de l'estimateur $\hat{\beta}$ par rapport à sa vrai valeur $\beta$.

$$\hat{\beta} - \beta = (X'X)^{-1}X'u$$

# Propriétés Estimateur MLE ET MCO

## Variance

On peut donc incorporer l'équation de $(\hat{\beta} - \beta)$ dans l'équation de la variance de $\hat{\beta}$

$$
\begin{aligned}
Var(\hat{\beta}) &= E[((X'X)^{-1}X'u)((X'X)^{-1}X'u)'] \\
&= E[(X'X)^{-1}X'uu'X(X'X)^{-1}] \\
&= (X'X)^{-1}X'E(uu')X(X'X)^{-1}
\end{aligned}
$$

Sachant $E(uu') = \sigma^2 I$

$$Var(\hat{\beta}) = (X'X)^{-1}X'\sigma^2 IX(X'X)^{-1}$$

Sachant $(X'X)^{-1}X'X) = I$

$$Var(\hat{\beta}) = \sigma^2 (X'X)^{-1}$$

**On voit donc que la variance de $\hat{\beta}$ atteint la borne de Crameur-Rao ou l'inverse de la matrice d'information**

# Propriétés de $\hat{\sigma}^2$

On sait que:

$$\frac{(Y - X\hat{\beta})'(Y - X\hat{\beta})}{\sigma^2} \sim X^2(T - K)$$

- Sachant $\hat{u}_t = Y - X\hat{\beta}$
  - $\hat{u}_t$ sont normales par hypothèses
  - $\hat{u}_t^{'}\hat{u}_t$ suit une loi chi carré

# Propriétés de $\hat{\sigma}^2$

**On peut montrer que l'espérance de ce terme est la suivante:**

$$E\left[\frac{(Y-X\beta)'(Y-X\beta)}{\sigma^2}\right]=(T-K)$$

**On peut montrer que la variance de ce terme est la suivante:**

$$V\left[\frac{(Y-X\beta)'(Y-X\beta)}{\sigma^2}\right]=2(T-K)$$

# Propriétés de $\hat{\sigma}^2$

**On sait que l'estimateur $\hat{\sigma}^2$ est le suivant:**

$$\hat{\sigma}^2 = \frac{(Y - X\beta)'(Y - X\beta)}{T}$$

**Espérance de $\hat{\sigma}^2$**

$$E(\hat{\sigma}^2) = E\left(\frac{(Y - X\beta)'(Y - X\beta)}{T}\right)$$

# Propriétés de $\hat{\sigma}^2$

- On peut multiplier le numérateur et le dénominateur par $\sigma^2$ afin d'écrire l'équation de l'espérance comme suit:

$$E(\hat{\sigma}^2) = E\left(\frac{(Y - X\beta)'(Y - X\beta)}{\sigma^2} \times \frac{\sigma^2}{T}\right)$$

- Sachant $E\left[\frac{(Y-X\beta)'(Y-X\beta)}{\sigma^2}\right] = (T - K)$ on peut formuler à l'équation de $\hat{\sigma}^2$ comme suit:

$$\frac{\sigma^2}{T}(T - K)$$

# Propriétés de $\hat{\sigma}^2$

- On voit que cette estimateur est biaisé et la borne de Crameur-Rao ne peu s'appliquer dans le ce cas.
- Cependant, si $T$ devient suffisament grand, alors:

$$T - K \approx T$$

- On voit clairement que le biais s'annule

$$E(\hat{\sigma}^2) = \frac{\sigma^2}{T}(T) = \sigma^2$$

# Estimateur de la variance sans biais

- L'estimateur de la variance sans biais est représenté par $\hat{S}^2$

$$\hat{S}^2 = \frac{(Y - X\hat{\beta})'(Y - X\hat{\beta})}{T - K}$$

**Espérance de** $\hat{S}^2$

$$E(\hat{S}^2) = E\left[\frac{(Y - X\hat{\beta})'(Y - X\hat{\beta})}{T - K}\right]$$

- On peut multiplier le numérateur et le dénominateur par $\sigma^2$ afin d'écrire l'équation de l'espérance de $\hat{S}^2$ comme suit:

$$E(\hat{S}^2) = E\left[\frac{(Y - X\hat{\beta})'(Y - X\hat{\beta})}{\sigma^2} \times \frac{\sigma^2}{T - K}\right]$$

# Estimateur de la variance sans biais

- Sachant $E\left[\frac{(Y-X\beta)'(Y-X\beta)}{\sigma^2}\right] = (T-K)$ on peut formuler à l'équation de l'espérance de $\hat{S}^2$ comme suit:

$$
\begin{aligned}
E(\hat{S}^2) &= (T-K) \times \frac{\sigma^2}{T-K} \\
&= \sigma^2
\end{aligned}
$$

- On voit donc que l'estimateur de la variance $\hat{S}^2$ est sans biais étant donnée que sont espérance égale la vrai valeur de la variance $\sigma^2$

# Estimateur de la variance sans biais

**Variance de $\hat{S}^2$**

$$V(\hat{S}^2) = V\left[\frac{(Y - X\hat{\beta})'(Y - X\hat{\beta})}{T - K}\right]$$

- On peut multiplier le numérateur et le dénominateur par $\sigma^2$ afin d'écrire l'équation de la variance de $\hat{S}^2$ comme suit:

$$V(\hat{S}^2) = V\left[\frac{(Y - X\hat{\beta})'(Y - X\hat{\beta})}{\sigma^2} \times \frac{\sigma^2}{T - K}\right]$$

- On peut sortir $\frac{\sigma^2}{T-K}$ de l'opérateur variance en élevant ce terme à la puissance 2.

$$V(\hat{S}^2) = \frac{\sigma^4}{(T - K)^2} V\left[\frac{(Y - X\hat{\beta})'(Y - X\hat{\beta})}{\sigma^2}\right]$$

# Estimateur de la variance sans biais

**Variance de $\hat{S}^2$**

- Sachant $V\left[\frac{(Y-X\beta)'(Y-X\beta)}{\sigma^2}\right] = 2(T-K)$, on peut écrire la variance de $\hat{S}^2$ comme suit:

$$\begin{aligned} V(\hat{S}^2) &= \frac{\sigma^4}{(T-K)^2} \times [2(T-K)] \\ &= \frac{2\sigma^4}{T-K} \end{aligned}$$

- On voit clairement que $V(\hat{S}^2)$ n'atteint pas la borne de Cramer-Rao étant donnée que cette variance est plus grande que celle donnée par la borne.

$$\frac{2\sigma^4}{T-K} > \frac{2\sigma^4}{T}$$