Le Maximum de Vraisemblance

GSF-6053

Hiver 2025

Table des matières

1 Introduction aux Moindres Carrés Ordinaires (MCO) 2

1.1 Objectif des MCO . . . 2

1.2 Hypothèses des MCO . . . 2

1.3 Propriétés des Estimateurs MCO . . . 3

1.4 Importance des MCO en Econométrie . . . 3

2 Le Maximum de Vraisemblance 4

2.1 Introduction . . . 4

2.2 Exemple : Modèle de Régression Linéaire . . . 5

2.2.1 La Vraisemblance Normale . . . 5

3 Propriété des Estimateurs du MLE et des MCO 7

3.1 Définitions . . . 7

3.2 Calculs des éléments de la matrice d'information . . . 9

3.3 Propriétés de $\hat{\beta}$ . . . 9

3.4 Propriétés de $\hat{\sigma}^2$ . . . 10

3.5 Variance de $\hat{s}^2$ . . . 11

4 Bibliographie 12

Bibliographie 12

1 Introduction aux Moindres Carrés Ordinaires (MCO)

Chapitres 3 de Gujarati et Porter, Chapitre 2 de Wooldridge.

Les Moindres Carrés Ordinaires (MCO) sont une méthode d'estimation utilisée en économétrie pour déterminer les paramètres d'un modèle de régression linéaire. L'objectif principal des MCO est de minimiser la somme des carrés des résidus, c'est-à-dire la somme des différences au carré entre les valeurs observées et les valeurs prédites par le modèle.

1.1 Objectif des MCO

L'objectif des MCO est d'estimer les paramètres β dans le modèle de régression linéaire suivant :

$$Y = X\beta + u,$$

où :

— Y est le vecteur des variables dépendantes.

— X est la matrice des variables explicatives (incluant une colonne de constantes si nécessaire).

— β est le vecteur des paramètres à estimer.

— u est le vecteur des termes d'erreur.

1.2 Hypothèses des MCO

Pour que les estimateurs des MCO soient les meilleurs estimateurs linéaires sans biais (BLUE - Best Linear Unbiased Estimators), plusieurs hypothèses doivent être respectées :

1. **Linéarité des Paramètres** : Le modèle doit être linéaire en ses paramètres β.
2. **Exogénéité des Régresseurs** : Les variables explicatives X doivent être exogènes, c'est-à-dire que $E(u \mid X) = 0$.

3. **Absence de Multicolinéarité Parfaite** : Les variables explicatives ne doivent pas être parfaitement corrélées entre elles.
4. **Homoscedasticité** : La variance des termes d'erreur u doit être constante, $\mathrm{Var}(u \mid X) = \sigma^2 I$.
5. **Indépendance des Termes d'Erreur** : Les termes d'erreur u doivent être indépendants entre eux.

1.3 Propriétés des Estimateurs MCO

Sous les hypothèses précédentes, les estimateurs des MCO $\hat{\beta}$ possèdent les propriétés suivantes :

— **Non Biaisés** : $E(\hat{\beta}) = \beta$.

— **Minimum de la Variance** : Parmi tous les estimateurs linéaires non biaisés, $\hat{\beta}$ a la variance la plus faible.

— **Consistance** : Lorsque la taille de l'échantillon augmente, $\hat{\beta}$ converge en probabilité vers β.

— **Normalité** : Si les termes d'erreur u sont normalement distribués, alors $\hat{\beta}$ suit une distribution normale.

1.4 Importance des MCO en Econométrie

Les MCO sont fondamentaux en économétrie car ils fournissent une méthode simple et efficace pour estimer les relations entre les variables économiques. Ils sont utilisés dans une multitude d'applications, telles que :

— Estimation de l'impact des facteurs économiques sur la consommation.

— Analyse de la relation entre l'investissement et le taux d'intérêt.

— Étude de l'effet des dépenses publicitaires sur les ventes.

Les MCO servent également de base pour des méthodes plus avancées en économétrie, telles que la régression avec variables endogènes, les modèles à effets fixes, et les modèles de panel.

2 Le Maximum de Vraisemblance

2.1 Introduction

Gujarati et Porter : Appendice 4.A et chapitre 7 (appendice 7.A), Appendice A à partir de la page 825.

Wooldridge appendice C C-4b.

Référence supplémentaire si nécessaire : Chapitre 14 dans le livre de Greene. En particulier 14.9.1 pour le cas du modèle de régression linéaire simple.

Soit θ le vecteur de paramètres à estimer. Le principe est le suivant :

Donner à θ la valeur qui maximise la probabilité d'obtenir un échantillon identique à celui qu'on dispose.

En détail, le principe est le suivant :

— La méthode du maximum de vraisemblance est une technique d'estimation des paramètres d'un modèle (d'une équation ou d'un système, linéaire ou non linéaire) avec ou sans restrictions sur les paramètres (coefficients, matrice de variances et covariances).

— Elle nécessite que la loi de distribution des observations soit connue (à des paramètres inconnus près).

— On construit une fonction des observations et des paramètres appelée fonction de vraisemblance construite à partir de la fonction de densité.

— On cherche les valeurs des paramètres qui maximisent cette fonction, étant données les observations dont on dispose.

— La valeur proposée comme estimateur des paramètres est celle qui rend le plus probable le fait que les observations que l'on a soient des réalisations de la loi de distribution ; cette valeur est appelée estimateur du maximum de vraisemblance (MLE).

Pour comprendre intuitivement l'estimateur de vraisemblance, regardons au tableau un exemple avec la distribution de Poisson en page 511 de Greene :

2.2 Exemple : Modèle de Régression Linéaire

2.2.1 La Vraisemblance Normale

Comme mentionné plus haut, le MLE nécessite de faire une hypothèse distributionnelle sur les u_t.

Sous l'hypothèse

$$u_t \sim \text{i.i.d. } N(0, \sigma^2)$$

La loi conjointe des $u_1, u_2, \ldots, u_T$ est le produit des normales de moyenne 0 et variance σ^2. Ainsi, par l'hypothèse d'indépendance, la loi conjointe des u_t est égale au produit des densités des u_t :

$$\begin{aligned} f(u_1, u_2, \ldots, u_T) &= \prod_{t=1}^{T} f(u_t) \\ &= \left(\sqrt{2\pi\sigma^2}\right)^{-T} \exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{T} u_t^2\right\} \end{aligned}$$

Nous sommes intéressés à la loi conjointe des Y_t. Effectuer un changement de variable : $u_t = Y_t - X_t'\beta$.

[1]

On peut écrire la loi conjointe des Y_t (la vraisemblance) :

$$\begin{aligned} g(Y_1, Y_2, \ldots, Y_T) &= \prod_{t=1}^{T} g(Y_t) \\ &= \left(\sqrt{2\pi\sigma^2}\right)^{-T} \exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{T} (Y_t - X_t'\beta)^2\right\} \\ &= (2\pi\sigma^2)^{-T/2} \exp\left\{-\frac{(Y - X\beta)'(Y - X\beta)}{2\sigma^2}\right\} \end{aligned}$$

1. Nous avons des hypothèses distributionnelles sur les erreurs, mais nous voulons écrire la vraisemblance des Y. Il faut donc passer des erreurs au Y avec un changement de variable. Si u est une variable aléatoire continue avec une pdf $f_x(x)$ et si $y = g(x)$ est une fonction continue et monotone de x.

Pour faciliter les calculs, on prend le logarithme de la fonction de vraisemblance. Elle transforme le produit en somme. Comme la transformation log est monotone, elle ne change pas le problème de maximisation. La log-vraisemblance (le log de la loi conjointe des Y_t) sera :

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

On veut donc maximiser cette fonction de vraisemblance en fonction de β et σ^2. Les conditions de premier ordre sont :

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \beta} = -\frac{1}{\sigma^2}\frac{\partial(Y'Y - 2Y'X\beta + \beta'X'X\beta)}{\partial \beta} = 0 \\ \Rightarrow \quad \hat{\beta} = (X'X)^{-1}X'Y\end{aligned}$$

$$\frac{\partial \mathcal{L}}{\partial \sigma^2} = -\frac{T}{\sigma^2} + \frac{(Y - X\beta)'(Y - X\beta)}{\sigma^4} = 0 \quad \Rightarrow \quad \hat{\sigma}^2 = \frac{(Y - X\beta)'(Y - X\beta)}{T}$$

Note : L'estimateur ci-haut est l'estimateur du maximum de vraisemblance. On voudra maintenant étudier ces propriétés. Nous allons démontrer que cet estimateur est sans biais et efficace sous certaines conditions. Nous allons également voir comment corriger le biais de l'estimateur de σ^2.

Le MLE de σ^2 est biaisé vers le bas. Ce n'est pas une propriété désirable. Par contre, sur base de notre démonstration, nous pourrons trouver une alternative sans biais :

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

C'est ce que nous étudions à présent.

3 Propriété des Estimateurs du MLE et des MCO

Chapitres 3 de Gujarati et Porter, Chapitre 3.5 et appendice D de Wooldridge.

On veut maintenant savoir si les estimateurs que nous avons trouvés sont bons. Il faut étudier le biais et la variance de l'estimateur.

Pour les MCO et l'estimateur de $\hat{\beta}$ du MLE qui coïncident, le résultat connu est le théorème de Gauss-Markov :

Théorème de Gauss-Markov : Étant donné les hypothèses du modèle de régression classique, les MCO sont des estimateurs sans biais qui ont la variance minimale. On dit qu'ils sont BLUE : Best Linear Unbiased Estimator.

La démonstration de ce théorème est effectuée dans l'appendice 3.A de Gujarati et Porter. Elle est aussi en section 3A.6 de Wooldridge. Ici, la notation matricielle et les estimateurs du maximum de vraisemblance me permettent de faire la démonstration d'une autre manière.

3.1 Définitions

Notre objectif pour démontrer que les MCO sont BLUE et de trouver des estimateurs

1. sans biais
2. efficaces.

Sans biais : L'espérance de l'estimateur égale à la vraie valeur du paramètre. (En moyenne, notre estimateur est égal au vrai paramètre qui nous intéresse.) Autrement dit :

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

Efficace : Un estimateur est efficace s'il atteint la borne de Cramér-Rao. On veut l'estimateur avec la variance la plus petite possible afin d'avoir une meilleure précision. (Ceci fait référence à la variance minimale de Gauss-Markov)

Plus tard, nous définirons aussi la propriété de convergence.

Borne de Cramér-Rao : Pour tout estimateur régulier et sans biais, sa variance est bornée par l'inverse de la matrice d'information.

La matrice d'information est quant à elle une façon de mesurer la quantité d'information sur les paramètres dans θ contenue dans X.

Une définition équivalente serait que la variance d'un estimateur sans biais sera toujours au moins aussi grande que l'inverse de la matrice d'information :

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

(See Appendix C of Greene for additional information on the Cramér-Rao bound p.1059). In our case,

$$\begin{aligned} \frac{\partial^2 \mathcal{L}}{\partial \beta \partial \beta'} &= -\frac{1}{\sigma^2}(X'X) \\ \frac{\partial^2 \mathcal{L}}{\partial \beta \partial \sigma^2} &= -\frac{1}{\sigma^4}(X'u) \\ \frac{\partial^2 \mathcal{L}}{\partial \sigma^2 \partial \beta'} &= -\frac{1}{\sigma^4}(u'X) \\ \frac{\partial^2 \mathcal{L}}{\partial (\sigma^2)^2} &= -\frac{T}{2\sigma^4} + \frac{(Y - X\beta)'(Y - X\beta)}{2\sigma^6} \end{aligned}$$

En prenant l'espérance mathématique de chaque dérivée, nous obtenons :

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

$$\mathbb{E}\left(-\frac{\partial^2 \mathcal{L}}{\partial(\sigma^2)^2}\right) = \frac{2\sigma^4}{T}$$

Ainsi, la matrice d'information est :

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

3.2 Calculs des éléments de la matrice d'information

$$\mathbb{E}\left[-\frac{\partial^2 \mathcal{L}}{\partial\beta\partial\beta'}\right] = \frac{1}{\sigma^2}(X'X)$$
$$\mathbb{E}\left[-\frac{\partial^2 \mathcal{L}}{\partial\beta\partial\sigma^2}\right] = 0$$
$$\mathbb{E}\left[-\frac{\partial^2 \mathcal{L}}{\partial(\sigma^2)^2}\right] = \frac{2\sigma^4}{T}$$

Ainsi,

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

3.3 Propriétés de $\hat{\beta}$

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

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

Donc, l'estimateur $\hat{\beta}$ des MCO et du MLE est sans biais. Regardons maintenant si sa variance atteint la borne minimale de Cramér-Rao (s'il est efficace).

$$\begin{aligned}V(\hat{\beta}) = \mathbb{E}[(\hat{\beta}-\beta)(\hat{\beta}-\beta)'] &= \mathbb{E}[(X'X)^{-1}X'uu'X(X'X)^{-1}] \\ &= (X'X)^{-1}X'\mathbb{E}(uu')X(X'X)^{-1} \\ &= \sigma^2(X'X)^{-1}\end{aligned}$$

Donc, $V(\hat{\beta})$ atteint la borne de Cramér-Rao.

3.4 Propriétés de $\hat{\sigma}^2$

$$\begin{aligned}\mathbb{E}(\hat{\sigma}^2) &= \mathbb{E}\left(\frac{(Y-X\hat{\beta})'(Y-X\hat{\beta})}{T}\right) \\ &= \frac{\sigma^2(T-K)}{T}\end{aligned}$$

Nous savons que :

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

Nous savons que les erreurs (u_t) sont normales par hypothèses, et les résidus $(\hat{u}_t)$ aussi par hypothèse. On sait donc que cette expression (qui est la somme des carrés des résidus) suit une loi chi carré (loi normale centrée élevée au carré).

De plus, on peut montrer que :

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

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

Cette expression n'est pas exactement égale à ce que l'on cherche soit $\hat{\sigma}^2$.

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

On voit donc que cette estimation est biaisée et la borne ne s'applique pas. Toutefois, lorsque T est suffisamment grand, $T - K \approx T$ et le biais s'annule.

Par contre, l'estimateur de la variance suivant sera sans biais :

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

Car :

$$\mathbb{E}(\hat{s}^2) = \sigma^2$$

[2]

3.5 Variance de $\hat{s}^2$

On peut maintenant vérifier si cet estimateur de la variance atteint la borne de Cramér-Rao. (Note : $\hat{s}^2 \neq \hat{\sigma}^2$, c'est un nouvel estimateur)

$$\begin{aligned} V(\hat{s}^2) =& \mathbb{E}\left[\left(\frac{(Y - X\hat{\beta})'(Y - X\hat{\beta})}{T-K}\right)^2\right] - \left(\mathbb{E}\left[\frac{(Y - X\hat{\beta})'(Y - X\hat{\beta})}{T-K}\right]\right)^2 \\ =& \frac{2\sigma^4}{T-K} \end{aligned}$$

Ce qui n'atteint pas la borne :

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

Encore une fois, si T est suffisamment grand et que K reste fixe, nous atteindrons presque la borne de variance minimale.

2. NB : si vous ne voulez pas faire cette hypothèse simplificatrice, vous pouvez regarder la preuve du théorème E.4 dans Wooldridge qui utilise plutôt la matrice de projection des X, M.

4 Bibliographie

Références

[1] Gujarati, D. N., & Porter, D. C. (2009). *Basic Econometrics*. McGraw-Hill.

[2] Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. MIT Press.

[3] Greene, W. H. (2012). *Econometric Analysis*. Pearson.