

Moindres Carrés Ordinaires

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	Régression Simple et Moindres Carrés Ordinaires : Approche en Sommation	4
2.1	Intuition de la Régression	4
2.2	La Régression : Introduction	5
2.3	Les Moindres Carrés Ordinaires (MCO)	6
3	Dérivation Matricielle des Moindres Carrés Ordinaires	7
3.1	Le Modèle	7
3.2	Hypothèses du Modèle de Régression	7
3.3	Dérivation des Estimateurs MCO	8
	Bibliographie	8

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 $E(u | X) = 0$.



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, $\text{Var}(u \mid X) = \sigma^2 I$.
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 Régression Simple et Moindres Carrés Ordinaires : Approche en Sommation

2.1 Intuition de la Régression

Chapitres 1 et 2 de [Gujarati et Porter](#), Chapitre 2 de [Wooldridge](#).

En termes généraux, l'analyse par régression s'intéresse à déterminer une relation de dépendance entre une variable dépendante et une ou plusieurs variables explicatives. L'objectif est d'estimer la valeur moyenne de la variable dépendante $E(Y)$ en termes de valeurs connues des variables explicatives ou indépendantes.

Il faut être prudent avec le vocabulaire employé. L'analyse par régression explique une dépendance, et non un lien de cause à effet. La *causalité* en économétrie est une relation plus forte de sens que la dépendance. Une relation statistique trouvée dans un échantillon de données n'implique pas nécessairement un lien de causalité. Bien qu'il existe des tests statistiques de causalité (tests de causalité de Granger ou de Sims), la causalité est souvent explicite dans le modèle théorique testé.

Il faut aussi différencier la régression de la corrélation. Le coefficient de corrélation mesure la force ou le degré d'une association linéaire entre deux variables. Par exemple, on peut vouloir mesurer le coefficient de corrélation entre un titre en particulier et un indice de référence. En revanche, la régression de ce titre sur un indice de référence estime la valeur moyenne du titre pour une variation dans l'indice. Aussi, la régression traite les variables de façon asymétrique : on suppose des variables explicatives *connues ou fixes*¹ alors que la variable dépendante est stochastique (avec une distribution de probabilité). En corrélation, les deux variables sont traitées de manière symétrique et sont toutes deux stochastiques.

1. Cette hypothèse peut être relâchée, mais les modèles deviennent plus complexes et les estimateurs ont alors des propriétés différentes du point de vue statistique.



2.2 La Régression : Introduction

On s'intéresse donc à trouver l'espérance conditionnelle d'une ou plusieurs variables dépendantes (Y) conditionnelle à la valeur connue des régresseurs (X) :

$$E(Y | X_i) = f(X_i), \quad (1.1)$$

où $f(X_i)$ est une fonction des variables explicatives X . Pour illustrer, supposons que nous avons une variable à expliquer (le rendement du titre Apple, Y) et une variable explicative (le rendement du marché, X_i). On assume que la relation entre Y et X est **linéaire** :

$$E(Y | X_i) = \beta_1 + \beta_2 X_i, \quad (1.2)$$

où β_1 (ordonnée à l'origine) et β_2 (pente) sont des paramètres fixes mais inconnus.

Puisque la régression explique la moyenne de Y , on s'attend à des déviations entre Y_i et $E(Y | X_i)$. Cette déviation ou *erreur* est :

$$u_i = Y_i - E(Y | X_i). \quad (1.3)$$

En pratique, le rendement d'Apple s'écrit alors :

$$Y_i = \beta_1 + \beta_2 X_i + u_i. \quad (1.4)$$

En prenant l'espérance conditionnelle, on obtient :

$$E(Y_i | X_i) = E(Y | X_i) + E(u_i).$$

Si $E(u_i) = 0$, alors $E(Y_i | X_i) = \beta_1 + \beta_2 X_i$.

Pourquoi ajouter un terme d'erreur ?

1. Le modèle théorique peut omettre certaines variables.
2. Certaines variables sont impossibles à mesurer (ex. : l'aversion au risque).



3. Il n'existe pas de relation certaine à 100%.
4. On privilégie des modèles parcimonieux ; les effets marginaux non spécifiés finissent dans l'erreur.
5. La forme fonctionnelle (par ex. linéaire) est une approximation.

Dans la pratique, on ne dispose que d'un échantillon de données, et non de la population entière. Les paramètres du modèle estimés à partir des données sont notés $\hat{\beta}_1$ et $\hat{\beta}_2$. Ainsi, le modèle estimé s'écrit :

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i, \quad (1.6)$$

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i. \quad (1.7)$$

Ici, $\hat{u}_i = Y_i - \hat{Y}_i$ est le *résidu*.

2.3 Les Moindres Carrés Ordinaires (MCO)

Chapitres 3 de [Gujarati et Porter](#).

Définissons d'abord le résidu :

$$\hat{u}_i = Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i.$$

L'idée des MCO est de choisir $\hat{\beta}_1$ et $\hat{\beta}_2$ de façon à minimiser la somme des carrés des résidus :

$$\sum_{i=1}^N \hat{u}_i^2 = \sum_{i=1}^N (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2.$$

En dérivant cette fonction (par rapport à $\hat{\beta}_1$ et $\hat{\beta}_2$) et en l'égalant à zéro, on obtient les conditions de premier ordre qui mènent aux *équations normales* :

$$\begin{aligned} \hat{\beta}_2 &= \frac{\sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^N (X_i - \bar{X})^2}, \\ \hat{\beta}_1 &= \bar{Y} - \hat{\beta}_2 \bar{X}. \end{aligned}$$



La droite de régression estimée passe par les moyennes \bar{X} et \bar{Y} . Dans la suite, on généralise cette méthode à plusieurs variables explicatives et on utilise la notation matricielle.

3 Dérivation Matricielle des Moindres Carrés Ordinaires

Chapitres 3 et 4, Appendice C de [Gujarati et Porter](#) ; Chapitres 2 et 3, Appendice E de [Wooldridge](#).

3.1 Le Modèle

Le modèle linéaire à K variables explicatives s'écrit :

$$Y = X\beta + u, \quad (2.1)$$

avec

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_T \end{bmatrix}, \quad X = \begin{bmatrix} 1 & X_{11} & \dots & X_{1,K-1} \\ 1 & X_{21} & \dots & X_{2,K-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{T1} & \dots & X_{T,K-1} \end{bmatrix}_{T \times K}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{K-1} \end{bmatrix}_{K \times 1}, \quad u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_T \end{bmatrix}_{T \times 1}.$$

La composante déterministe est $X\beta$, la composante aléatoire est u . L'hypothèse clé est que $E(u | X) = 0$ et que les composantes de u sont supposées iid, sans corrélation ni hétéroscédasticité.

3.2 Hypothèses du Modèle de Régression

1. Linéarité des paramètres.
2. Les régresseurs X sont fixes (ou exogènes).
3. $T > K$ et $\text{rang}(X) = K$.
4. Pas de multicollinéarité parfaite.



3.3 Dérivation des Estimateurs MCO

La méthode consiste à minimiser la somme des carrés des résidus :

$$\hat{\beta} = \arg \min_{\beta} (Y - X\beta)'(Y - X\beta).$$

Par dérivation et condition de premier ordre, on obtient :

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

La variance du terme d'erreur peut être estimée par :

$$\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^T \hat{u}_t^2 = \frac{\hat{u}'\hat{u}}{T},$$

où $\hat{u} = Y - X\hat{\beta}$.

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.

