

UNIVERSITÉ DE GRENOBLE ALPES  
(L3 MIASH, S2)

## **ÉCONOMÉTRIE**

### **RÉGRESSION LINÉAIRE ET MOINDRES CARRÉS À DISTANCE FINIE(2)**

#### **ESTIMATEUR DES MCO**

(CETTE VERSION : 11 FÉVRIER 2024)

MICHAL W. URDANIVIA <sup>1</sup>

---

1. Contact : [michal.wong-urdanivia@univ-grenoble-alpes.fr](mailto:michal.wong-urdanivia@univ-grenoble-alpes.fr), Université de Grenoble Alpes, Faculté d'Économie, GAEL.

## TABLE DES MATIÈRES

1. Matrices de projection	2
2. Propriétés de $\hat{\sigma}^2$	4
3. Régression partitionnée	5
4. Qualité de l'ajustement et coefficient de détermination ou $R^2$	8
5. Propriétés du $R^2$	9
6. $R^2$ ajusté	11

## 1. MATRICES DE PROJECTION

Nous pouvons penser à  $\mathbf{Y}$  et aux colonnes  $\mathbf{X}$  comme des éléments de l'espace euclidien à  $n$  dimensions,  $\mathbb{R}^n$ . On peut définir un sous-espace de  $\mathbb{R}^n$  appelé l'*espace des colonnes* d'une matrice  $n \times K$ ,  $\mathbf{X}$ . Il s'agit de la collection de tous les vecteurs dans  $\mathbb{R}^n$  qui peuvent s'écrire comme des combinaisons linéaires des colonnes de  $\mathbf{X}$ ,

$$\mathcal{S}(\mathbf{X}) = \{z \in \mathbb{R}^n : z = \mathbf{X}b, b = (b_1, b_2, \dots, b_K) \in \mathbb{R}^K\}$$

Étant donné deux vecteurs  $a, b$ , dans  $\mathbb{R}^n$ , la distance entre  $a$  et  $b$  est donné par la norme euclidienne<sup>2</sup> de leur différence  $\|a - b\| = \sqrt{(a - b)^\top (a - b)}$ . En conséquence, le problème des moindres carrés, à savoir la minimisation de la somme des carrés des erreurs,  $(\mathbf{Y} - \mathbf{X}b)^\top (\mathbf{Y} - \mathbf{X}b)$ , est celui de trouver, parmi tous les éléments de  $\mathcal{S}(\mathbf{X})$ , celui dont la distance par rapport à  $\mathbf{Y}$  est la plus petite,

$$\min_{\hat{\mathbf{Y}} \in \mathcal{S}(\mathbf{X})} \|\mathbf{Y} - \hat{\mathbf{Y}}\|^2$$

Le point le plus proche est obtenu en "traçant une perpendiculaire". Autrement dit, une solution au problème des moindres carrés,  $\hat{\mathbf{Y}} = \mathbf{X}\hat{\beta}$  doit être choisie de sorte que le vecteur des résidus,  $\hat{\mathbf{U}} = \mathbf{Y} - \hat{\mathbf{Y}}$  soit orthogonal(perpendiculaire) à chaque colonne de  $\mathbf{X}$ ,

$$\hat{\mathbf{U}}^\top \mathbf{X} = 0$$

Un résultat de cela est que  $\hat{\mathbf{U}}$  est orthogonal à chaque élément de  $\mathcal{S}(\mathbf{X})$ . En effet, si  $z \in \mathcal{S}(\mathbf{X})$ , alors il existe  $b \in \mathbb{R}^K$  tel que  $z = \mathbf{X}b$ , et,

$$\begin{aligned} \hat{\mathbf{U}}^\top z &= \hat{\mathbf{U}}^\top \mathbf{X}b \\ &= 0 \end{aligned}$$

La collection des éléments de  $\mathbb{R}^n$  orthogonaux à  $\mathcal{S}(\mathbf{X})$  est appelée *complément orthogonal* de  $\mathcal{S}(\mathbf{X})$ ,

$$\mathcal{S}^\perp(\mathbf{X}) = \{z \in \mathbb{R}^n : z^\top \mathbf{X} = 0\}$$

Soulignons que tout élément de  $\mathcal{S}^\perp(\mathbf{X})$  est orthogonal à chaque élément de  $\mathcal{S}(\mathbf{X})$ .

Comme nous l'avons vu dans le cours précédent, la solution au problème des moindres carrés est donnée par,

$$\begin{aligned} \hat{\mathbf{Y}} &= \mathbf{X}\hat{\beta} \\ &= \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y} \\ &= \mathbf{P}_\mathbf{X} \mathbf{Y} \end{aligned}$$

où

$$\mathbf{P}_\mathbf{X} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$$

est appelée *matrice de projection orthogonale*. Pour tout vecteur  $\mathbf{Y} \in \mathbb{R}^n$ ,

$$\mathbf{P}_\mathbf{X} \mathbf{Y} \in \mathcal{S}(\mathbf{X})$$

En outre, le vecteur des résidus est dans  $\mathcal{S}^\perp(\mathbf{X})$ ,

$$\mathbf{Y} - \mathbf{P}_\mathbf{X} \mathbf{Y} \in \mathcal{S}^\perp(\mathbf{X}) \tag{1}$$

Pour montrer (1), notons d'abord, qu'étant donné que les colonnes de  $\mathbf{X}$  sont dans  $\mathcal{S}(\mathbf{X})$ ,

$$\begin{aligned} \mathbf{P}_\mathbf{X} \mathbf{X} &= \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{X} \\ &= \mathbf{X} \end{aligned}$$

2. Pour un vecteur  $x = (x_1, x_2, \dots, x_n)$  sa norme euclidienne est définie comme  $\|x\| = \sqrt{x^\top x} = \sqrt{\sum_{i=1}^n x_i^2}$ .

et comme  $\mathbf{P}_X$  est une matrice symétrique,

$$\mathbf{X}^\top \mathbf{P}_X = \mathbf{X}^\top$$

Maintenant,

$$\begin{aligned} \mathbf{X}^\top (\mathbf{Y} - \mathbf{P}_X \mathbf{Y}) &= \mathbf{X}^\top \mathbf{Y} - \mathbf{X}^\top \mathbf{P}_X \mathbf{Y} \\ &= \mathbf{X}^\top \mathbf{Y} - \mathbf{X}^\top \mathbf{Y} \\ &= 0 \end{aligned}$$

Ainsi, par définition, les résidus  $\mathbf{Y} - \mathbf{P}_X \mathbf{Y} \in \mathcal{S}^\perp(\mathbf{X})$ . Les résidus peuvent s'écrire,

$$\begin{aligned} \hat{\mathbf{U}} &= \mathbf{Y} - \mathbf{P}_X \mathbf{Y} \\ &= (\mathbf{I}_n - \mathbf{P}_X) \mathbf{Y} \\ &= \mathbf{M}_X \mathbf{Y} \end{aligned}$$

où,

$$\begin{aligned} \mathbf{M}_X &= \mathbf{I}_n - \mathbf{P}_X \\ &= \mathbf{I}_n - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \end{aligned}$$

est une matrice de projection dans  $\mathcal{S}^\perp(\mathbf{X})$ .

Les matrices  $\mathbf{P}_X$  et  $\mathbf{M}_X$  présentent les propriétés suivantes.

- (1)  $\mathbf{P}_X + \mathbf{M}_X = \mathbf{I}_n$ . Ceci implique, que pour tout  $\mathbf{Y} \in \mathbb{R}^n$ ,

$$\mathbf{Y} = \mathbf{P}_X \mathbf{Y} + \mathbf{M}_X \mathbf{Y}$$

- (2)  $\mathbf{P}_X$  et  $\mathbf{M}_X$  sont symétriques,

$$\mathbf{P}_X^\top = \mathbf{P}_X, \quad \mathbf{M}_X^\top = \mathbf{M}_X$$

- (3)  $\mathbf{P}_X$  et  $\mathbf{M}_X$  sont idempotentes,

$$\mathbf{P}_X \mathbf{P}_X = \mathbf{P}_X, \quad \mathbf{M}_X \mathbf{M}_X = \mathbf{M}_X$$

En effet,

$$\begin{aligned} \mathbf{P}_X \mathbf{P}_X &= (\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top) (\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top) \\ &= \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \\ &= \mathbf{P}_X \end{aligned}$$

de même,

$$\begin{aligned} \mathbf{M}_X \mathbf{M}_X &= (\mathbf{I}_n - \mathbf{P}_X)(\mathbf{I}_n - \mathbf{P}_X) \\ &= \mathbf{I}_n - 2\mathbf{P}_X + \mathbf{P}_X \mathbf{P}_X \\ &= \mathbf{I}_n - \mathbf{P}_X \\ &= \mathbf{M}_X \end{aligned}$$

- (4)  $\mathbf{P}_X$  et  $\mathbf{M}_X$  sont orthogonales,

$$\begin{aligned} \mathbf{P}_X \mathbf{M}_X &= \mathbf{P}_X (\mathbf{I}_n - \mathbf{P}_X) \\ &= \mathbf{P}_X - \mathbf{P}_X \mathbf{P}_X \\ &= \mathbf{P}_X - \mathbf{P}_X \\ &= 0 \end{aligned}$$

Cette propriété implique que  $\mathbf{M}_X \mathbf{X} = 0$ . En effet,

$$\begin{aligned}\mathbf{M}_X \mathbf{X} &= (\mathbf{I}_n - \mathbf{P}_X) \mathbf{X} \\ &= \mathbf{X} - \mathbf{P}_X \mathbf{X} \\ &= \mathbf{X} - \mathbf{X} \\ &= 0\end{aligned}$$

Observons que, dans la discussion ci-dessus, aucune des hypothèses quant au modèle de régression n'ont été utilisées. Étant donné des données,  $\mathbf{Y}$  et  $\mathbf{X}$ , nous pouvons toujours calculer l'estimateur des moindres carrés, indépendamment du processus générateur des données derrière les données. Néanmoins, nous avons besoin d'un modèle (i.e., d'hypothèses) pour pouvoir discuter des propriétés d'un estimateur (e.g., le fait qu'il soit ou non sans biais, etc).

## 2. PROPRIÉTÉS DE $\hat{\sigma}^2$

Nous avons suggéré précédemment d'estimer  $\sigma^2$  par,

$$\begin{aligned}\hat{\sigma}^2 &= n^{-1} \sum_{i=1}^n (Y_i - X_i^\top \hat{\beta})^2 \\ &= n^{-1} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \\ &= n^{-1} \sum_{i=1}^n \hat{U}_i^2 \\ &= n^{-1} \hat{\mathbf{U}}^\top \hat{\mathbf{U}}\end{aligned}$$

Il s'avère cependant que sous les hypothèses usuelles,  $(??) - (??)$ ,  $\hat{\sigma}^2$  est un estimateur biaisé. Pour le voir, écrivons d'abord,

$$\begin{aligned}\hat{\mathbf{U}} &= \mathbf{M}_X \mathbf{Y} \\ &= \mathbf{M}_X (\mathbf{X}\beta + \mathbf{U}) \\ &= \mathbf{M}_X \mathbf{U}\end{aligned}$$

où la dernière égalité résulte de ce que  $\mathbf{M}_X \mathbf{X} = 0$ . A présent,

$$\begin{aligned}n\hat{\sigma}^2 &= \hat{\mathbf{U}}^\top \hat{\mathbf{U}} \\ &= \mathbf{U}^\top \mathbf{M}_X \mathbf{M}_X \mathbf{U} \\ &= \mathbf{U}^\top \mathbf{M}_X \mathbf{U}\end{aligned}$$

Étant donné que  $\mathbf{U}^\top \mathbf{M}_X \mathbf{U}$  est un scalaire,

$$\mathbf{U}^\top \mathbf{M}_X \mathbf{U} = \text{Tr}(\mathbf{U}^\top \mathbf{M}_X \mathbf{U})$$

où  $\text{Tr}(A)$  désigne la trace de la matrice  $A$ . Nous avons,

$$\begin{aligned}\mathbb{E}(\mathbf{U}^\top \mathbf{M}_X \mathbf{U} | \mathbf{X}) &= \mathbb{E}(\text{Tr}(\mathbf{U}^\top \mathbf{M}_X \mathbf{U}) | \mathbf{X}) \\ &= \mathbb{E}(\text{Tr}(\mathbf{M}_X \mathbf{U} \mathbf{U}^\top) | \mathbf{X}) \quad (\text{car } \text{Tr}(ABC) = \text{Tr}(BCA)) \\ &= \text{Tr}(\mathbf{M}_X \mathbb{E}(\mathbf{U} \mathbf{U}^\top) | \mathbf{X}) \quad (\text{car l'opérateur trace et l'espérance sont linéaires}) \\ &= \sigma^2 \text{Tr}(\mathbf{M}_X)\end{aligned}$$

La dernière égalité résulte de ce que par l'hypothèse (??),  $E(\mathbf{U}^\top \mathbf{U}) = \sigma^2 \mathbf{I}_n$ . Maintenant,

$$\begin{aligned} \text{Tr}(\mathbf{M}_\mathbf{X}) &= \text{Tr}(\mathbf{I}_n - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top) \\ &= \text{Tr}(\mathbf{I}_n) - \text{Tr}(\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top) \\ &= \text{Tr}(\mathbf{I}_n) - \text{Tr}((\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{X}) \\ &= \text{Tr}(\mathbf{I}_n) - \text{Tr}(\mathbf{I}_K) \\ &= n - K \end{aligned}$$

Il s'en suit que,

$$E(\hat{\sigma}^2) = \frac{n-K}{n} \sigma^2 \quad (2)$$

L'estimateur  $\hat{\sigma}^2$  est biaisé, mais le résultat précédent suggère qu'il est aisé de le modifier afin d'obtenir un estimateur sans biais. Pour cela, définissons,

$$\begin{aligned} s^2 &= \hat{\sigma}^2 \frac{n}{n-K} \\ &= (n-K)^{-1} \sum_{i=1}^n \hat{U}_i^2 \end{aligned}$$

il résulte de (2) que,

$$E(s^2) = \sigma^2$$

### 3. RÉGRESSION PARTITIONNÉE

Considérons la partition de la matrice des régresseurs,  $\mathbf{X}$ ,

$$\mathbf{X} = (\mathbf{X}_1 \mathbf{X}_2)$$

et écrivons le modèle comme suit,

$$\mathbf{Y} = \mathbf{X}_1 \beta_1 + \mathbf{X}_2 \beta_2 + \mathbf{U}$$

où  $\mathbf{X}_1$  est une matrice  $(n \times K_1)$ ,  $\mathbf{X}_2$  est une matrice  $(n \times K_2)$ ,  $K_1 + K_2 = K$ , et,

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$$

$\beta_1$  et  $\beta_2$  étant des vecteurs de paramètres, respectivement,  $(K_1 \times 1)$  et  $(K_2 \times 1)$ . Partant de cette décomposition du modèle de régression concentrons nous sur un groupe de variables et leurs paramètres correspondants, par exemple  $\mathbf{X}_1$  et  $\beta_1$ .

Soit l'estimateur des moindres carrés de  $\beta$ ,

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix}$$

Nous pouvons écrire la version suivante des équations normales,

$$(\mathbf{X}^\top \mathbf{X}) \hat{\beta} = \mathbf{X}^\top \mathbf{Y}$$

comme suit,

$$\begin{pmatrix} \mathbf{X}_1^\top \mathbf{X}_1 & \mathbf{X}_1^\top \mathbf{X}_2 \\ \mathbf{X}_2^\top \mathbf{X}_1 & \mathbf{X}_2^\top \mathbf{X}_2 \end{pmatrix} \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1^\top \mathbf{Y} \\ \mathbf{X}_2^\top \mathbf{Y} \end{pmatrix}$$

On peut obtenir des expressions pour  $\hat{\beta}_1$  et  $\hat{\beta}_2$  par inversion de la matrice partitionnée à gauche de l'équation ci-dessus. Alternativement, définissons  $\mathbf{M}_2$  comme la matrice de projection sur l'espace orthogonal à l'espace  $\mathcal{S}(\mathbf{X}_2)$ ,

$$\mathbf{M}_2 = \mathbf{I}_n - \mathbf{X}_2(\mathbf{X}_2^\top \mathbf{X}_2)^{-1} \mathbf{X}_2^\top$$

alors,

$$\hat{\beta}_1 = (\mathbf{X}_1^\top \mathbf{M}_2 \mathbf{X}_1)^{-1} \mathbf{X}_1^\top \mathbf{M}_2 \mathbf{Y} \quad (3)$$

Pour montrer cela, commençons par écrire,

$$\mathbf{Y} = \mathbf{X}_1 \hat{\beta}_1 + \mathbf{X}_2 \hat{\beta}_2 + \hat{\mathbf{U}} \quad (4)$$

Notons que par construction,

$$\mathbf{M}_2 \hat{\mathbf{U}} = \hat{\mathbf{U}} (\hat{\mathbf{U}} \text{ est orthogonal à } \mathbf{X}_2)$$

$$\mathbf{M}_2 \mathbf{X}_2 = 0$$

$$\mathbf{X}_1^\top \hat{\mathbf{U}} = 0$$

$$\mathbf{X}_2^\top \hat{\mathbf{U}} = 0$$

Substituons l'équation (4) dans la partie droite de l'équation (3),

$$\begin{aligned} (\mathbf{X}_1^\top \mathbf{M}_2 \mathbf{X}_1)^{-1} \mathbf{X}_1^\top \mathbf{M}_2 \mathbf{Y} &= (\mathbf{X}_1^\top \mathbf{M}_2 \mathbf{X}_1)^{-1} \mathbf{X}_1^\top \mathbf{M}_2 (\mathbf{X}_1 \hat{\beta}_1 + \mathbf{X}_2 \hat{\beta}_2 + \hat{\mathbf{U}}) \\ &= (\mathbf{X}_1^\top \mathbf{M}_2 \mathbf{X}_1)^{-1} \mathbf{X}_1^\top \mathbf{M}_2 \mathbf{X}_1 \hat{\beta}_1 \\ &\quad + (\mathbf{X}_1^\top \mathbf{M}_2 \mathbf{X}_1)^{-1} \mathbf{X}_1^\top \hat{\mathbf{U}} \quad (\text{car } \mathbf{M}_2 \mathbf{X}_2 = 0 \text{ et } \mathbf{M}_2 \hat{\mathbf{U}} = \hat{\mathbf{U}}) \\ &= \hat{\beta}_1 \end{aligned}$$

Étant donné que  $\mathbf{M}_2$  est symétrique et idempotente, on peut écrire,

$$\begin{aligned} \hat{\beta}_1 &= ((\mathbf{M}_2 \mathbf{X}_1)^\top (\mathbf{M}_2 \mathbf{X}_1))^{-1} (\mathbf{M}_2 \mathbf{X}_1)^\top (\mathbf{M}_2 \mathbf{Y}) \\ &= (\tilde{\mathbf{X}}_1^\top \tilde{\mathbf{X}}_1)^{-1} \tilde{\mathbf{X}}_1^\top \tilde{\mathbf{Y}} \end{aligned}$$

où,

$$\begin{aligned} \tilde{\mathbf{X}}_1 &= \mathbf{M}_2 \mathbf{X}_1 \\ &= \mathbf{X}_1 - \mathbf{X}_2 (\mathbf{X}_2^\top \mathbf{X}_2)^{-1} \mathbf{X}_2^\top \mathbf{X}_1 \end{aligned}$$

à savoir les résidus de la régression de  $\mathbf{X}_1$  sur  $\mathbf{X}_2$ . Et où,

$$\begin{aligned} \tilde{\mathbf{Y}} &= \mathbf{M}_2 \mathbf{Y} \\ &= \mathbf{Y} - \mathbf{X}_2 (\mathbf{X}_2^\top \mathbf{X}_2)^{-1} \mathbf{X}_2^\top \mathbf{Y} \end{aligned}$$

à savoir les résidus de la régression de  $\mathbf{Y}$  sur  $\mathbf{X}_2$ .

Ainsi, pour obtenir les coefficients de  $K_1$  premiers régresseurs, plutôt que de réaliser la régression avec les  $K_1 + K_2 = K$  régresseurs, on peut régresser  $\mathbf{Y}$  sur  $\mathbf{X}_2$  pour obtenir les résidus  $\tilde{\mathbf{Y}}$ , régresser  $\mathbf{X}_1$  sur  $\mathbf{X}_2$  pour obtenir les résidus  $\tilde{\mathbf{X}}_1$ , et alors régresser  $\tilde{\mathbf{Y}}$  sur  $\tilde{\mathbf{X}}_1$  pour obtenir  $\hat{\beta}_1$ . Autrement dit,  $\hat{\beta}_1$  décrit l'effet de  $\mathbf{X}_1$  une fois que ceux de  $\mathbf{X}_2$  ont été contrôlés.

De manière similaire que pour  $\hat{\beta}_1$ , nous avons pour  $\hat{\beta}_2$ ,

$$\hat{\beta}_2 = (\mathbf{X}_2^\top \mathbf{M}_1 \mathbf{X}_2)^{-1} \mathbf{X}_2^\top \mathbf{M}_1 \mathbf{Y}$$

où,

$$\mathbf{M}_1 = \mathbf{I}_n - \mathbf{X}_1(\mathbf{X}_1^\top \mathbf{X}_1)^{-1} \mathbf{X}_1^\top$$

Prenons comme exemple le modèle suivant,

$$Y_i = \beta_1 + \beta_2 X_i + U_i, \quad i = 1, 2, \dots, n$$

Soit  $\mathbf{1}_n$  le vecteur  $(n \times 1)$  dont tous les éléments sont le nombre 1, i.e.,

$$\mathbf{1}_n = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ \vdots \\ 1 \end{pmatrix}$$

La matrice des régresseurs est alors,

$$(\mathbf{1}_n \ X) = \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ \vdots & \vdots \\ 1 & X_n \end{pmatrix}$$

Considérons,

$$\mathbf{M}_1 = \mathbf{I}_n - \mathbf{1}_n(\mathbf{1}_n^\top \mathbf{1}_n)^{-1} \mathbf{1}_n^\top$$

et,

$$\hat{\beta}_2 = \frac{X^\top \mathbf{M}_1 \mathbf{Y}}{X^\top \mathbf{M}_1 X}$$

Nous avons,  $\mathbf{1}_n^\top \mathbf{1}_n = n$ , par conséquent,

$$\mathbf{M}_1 = \mathbf{I}_n - \frac{\mathbf{1}_n \mathbf{1}_n^\top}{n}$$

et,

$$\begin{aligned} \mathbf{M}_1 X &= X - \mathbf{1}_n \frac{\mathbf{1}_n^\top X}{n} \\ &= X - \bar{X} \mathbf{1}_n \\ &= \begin{pmatrix} X_1 - \bar{X} \\ X_2 - \bar{X} \\ \vdots \\ \vdots \\ X_n - \bar{X} \end{pmatrix} \end{aligned}$$

où,

$$\begin{aligned} \bar{X} &= \frac{\mathbf{1}_n^\top X}{n} \\ &= n^{-1} \sum_{i=1}^n X_i \end{aligned}$$



Ainsi la matrice  $\mathbf{M}_1$  transforme le vecteur  $X$  en un vecteur dont les éléments sont les écarts des observations  $X_i$  à leur moyenne. Et nous pouvons écrire,

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

#### 4. QUALITÉ DE L' AJUSTEMENT ET COEFFICIENT DE DÉTERMINATION OU $R^2$

Écrivons,

$$\begin{aligned}\mathbf{Y} &= \mathbf{P}_X \mathbf{Y} + \mathbf{M}_X \mathbf{Y} \\ &= \hat{\mathbf{Y}} + \hat{\mathbf{U}}\end{aligned}$$

où par construction,

$$\begin{aligned}\hat{\mathbf{Y}}^\top \hat{\mathbf{U}} &= (\mathbf{P}_X \mathbf{Y})^\top (\mathbf{M}_X \mathbf{Y}) \\ &= \mathbf{Y}^\top \mathbf{P}_X \mathbf{M}_X \mathbf{Y} \\ &= 0\end{aligned}$$

Supposons que le modèle contienne une constante, par exemple la première colonne de la matrice des régresseurs  $\mathbf{X}$  est le vecteur unitaire  $\mathbf{1}_n$ . La *variation totale* dans  $\mathbf{Y}$  est,

$$\begin{aligned}\sum_{i=1}^n (Y_i - \bar{Y})^2 &= \mathbf{Y}^\top \mathbf{M}_1 \mathbf{Y} \\ &= (\hat{\mathbf{Y}} + \hat{\mathbf{U}})^\top \mathbf{M}_1 (\hat{\mathbf{Y}} + \hat{\mathbf{U}}) \\ &= \hat{\mathbf{Y}}^\top \mathbf{M}_1 \hat{\mathbf{Y}} + \hat{\mathbf{U}}^\top \mathbf{M}_1 \hat{\mathbf{U}} + 2\hat{\mathbf{Y}}^\top \mathbf{M}_1 \hat{\mathbf{U}}\end{aligned}$$

où  $\bar{Y} = n^{-1} \sum_{i=1}^n Y_i$ . Comme le modèle contient une constante,

$$\mathbf{1}_n^\top \hat{\mathbf{U}} = 0$$

et,

$$\mathbf{M}_1 \hat{\mathbf{U}} = \hat{\mathbf{U}}$$

Cependant,  $\hat{\mathbf{Y}}^\top \hat{\mathbf{U}} = 0$ , et par conséquent,

$$\mathbf{Y}^\top \mathbf{M}_1 \mathbf{Y} = \hat{\mathbf{Y}}^\top \mathbf{M}_1 \hat{\mathbf{Y}} + \hat{\mathbf{U}}^\top \hat{\mathbf{U}}$$

ou,

$$\sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n (\hat{Y}_i - \bar{\hat{Y}})^2 + \sum_{i=1}^n \hat{U}_i^2$$

où  $\bar{\hat{Y}} = n^{-1} \sum_{i=1}^n \hat{Y}_i$ . Notons que,

$$\begin{aligned}\bar{Y} &= \frac{\mathbf{1}_n^\top \mathbf{Y}}{n} \\ &= \frac{\mathbf{1}_n^\top \hat{\mathbf{Y}}}{n} + \frac{\mathbf{1}_n^\top \hat{\mathbf{U}}}{n} \\ &= \frac{\mathbf{1}_n^\top \hat{\mathbf{Y}}}{n} \\ &= \bar{\hat{Y}}\end{aligned}$$

Ainsi, la moyenne des  $Y_i$  et celle de leurs valeurs ajustées  $\hat{Y}_i$  étant égales, nous pouvons écrire,

$$\sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^n \hat{U}_i^2$$

ou,

$$SCT = SCE + SCR$$

où,  $SCT := \sum_{i=1}^n (Y_i - \bar{Y})^2$  est la *somme des carrés totale*,  $SCE := \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$  est la *somme des carrés expliqués*, et  $SCR := \sum_{i=1}^n \hat{U}_i^2$  est la *somme des carrés des résidus*.

Le rapport de la  $SCE$  à la  $SCT$  est appelé coefficient de détermination<sup>3</sup> ou  $R^2$ ,

$$\begin{aligned}R^2 &= \frac{SCE}{SCT} \\ &= \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \\ &= 1 - \frac{\sum_{i=1}^n \hat{U}_i^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \\ &= 1 - \frac{\hat{\mathbf{U}}^\top \hat{\mathbf{U}}}{\mathbf{Y}^\top \mathbf{M}_1 \mathbf{Y}}\end{aligned}$$

## 5. PROPRIÉTÉS DU $R^2$

- (1) Le  $R^2$  est borné entre 0 et 1 ainsi que cela est indiqué par sa décomposition. Remarquez néanmoins que ceci n'est plus vrai dans un modèle sans constante, et dans ce cas il est indiqué de ne pas utiliser la définition précédente du  $R^2$ . Remarquez aussi que si  $R^2 = 1$  alors  $\hat{\mathbf{U}}^\top \hat{\mathbf{U}} = 0$ , ce qui sera vrai seulement si  $\mathbf{Y} \in \mathcal{S}(X)$ , i.e.,  $\mathbf{Y}$  est *exactement* une combinaison linéaire des colonnes de  $\mathbf{X}$ .
- (2) Le  $R^2$  augmente avec le nombre de régresseurs. Pour montrer cette propriété considérons une partition de la matrice des régresseurs  $\mathbf{X} = (\mathbf{Z} \ \mathbf{W})$ . Étudions l'effet d'ajouter  $\mathbf{W}$  sur le  $R^2$ . Notons,

$$\begin{aligned}\mathbf{P}_X &= \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \\ \mathbf{P}_Z &= \mathbf{Z}(\mathbf{Z}^\top \mathbf{Z})^{-1} \mathbf{Z}^\top\end{aligned}$$

3. On l'appelle/prononce généralement "R deux".

respectivement, la matrice de projection du modèle "complet" (i.e., avec  $\mathbf{Z}$  et  $\mathbf{W}$ ), et la matrice de projection du modèle avec uniquement  $\mathbf{Z}$  comme matrice des régresseurs. Définissons aussi,

$$\begin{aligned}\mathbf{M}_X &= \mathbf{I}_n - \mathbf{P}_X \\ \mathbf{M}_Z &= \mathbf{I}_n - \mathbf{P}_Z\end{aligned}$$

Observons que comme  $\mathbf{Z}$  est une partie de  $\mathbf{X}$ ,

$$\mathbf{P}_X \mathbf{Z} = \mathbf{Z}$$

et,

$$\begin{aligned}\mathbf{P}_X \mathbf{P}_Z &= \mathbf{P}_X \mathbf{Z} (\mathbf{Z}^\top \mathbf{Z})^{-1} \mathbf{Z}^\top \\ &= \mathbf{Z} (\mathbf{Z}^\top \mathbf{Z})^{-1} \mathbf{Z}^\top \\ &= \mathbf{P}_Z\end{aligned}$$

Par conséquent,

$$\begin{aligned}\mathbf{M}_X \mathbf{M}_Z &= (\mathbf{I}_n - \mathbf{P}_X)(\mathbf{I}_n - \mathbf{P}_Z) \\ &= \mathbf{I}_n - \mathbf{P}_X - \mathbf{P}_Z + \mathbf{P}_X \mathbf{P}_Z \\ &= \mathbf{I}_n - \mathbf{P}_X - \mathbf{P}_Z + \mathbf{P}_Z \\ &= \mathbf{M}_X\end{aligned}$$

Supposons que  $\mathbf{Z}$  contienne un vecteur constant de sorte que les deux régressions (la complète et celle sans  $\mathbf{W}$ ) contiennent chacune une constante. Définissons,

$$\hat{\mathbf{U}}_X = \mathbf{M}_X \mathbf{Y}, \quad \hat{\mathbf{U}}_Z = \mathbf{M}_Z \mathbf{Y}$$

Écrivons,

$$(\hat{\mathbf{U}}_X - \hat{\mathbf{U}}_Z)^\top (\hat{\mathbf{U}}_X - \hat{\mathbf{U}}_Z) = \hat{\mathbf{U}}_X^\top \hat{\mathbf{U}}_X + \hat{\mathbf{U}}_Z^\top \hat{\mathbf{U}}_Z - 2\hat{\mathbf{U}}_X^\top \hat{\mathbf{U}}_Z \geq 0$$

Notons que,

$$\begin{aligned}\hat{\mathbf{U}}_X^\top \hat{\mathbf{U}}_Z &= \mathbf{Y}^\top \mathbf{M}_X \mathbf{M}_Z \mathbf{Y} \\ &= \mathbf{Y}^\top \mathbf{M}_X \mathbf{Y} \\ &= \hat{\mathbf{U}}_X^\top \hat{\mathbf{U}}_X\end{aligned}$$

d'où,

$$\hat{\mathbf{U}}_Z^\top \hat{\mathbf{U}}_Z \geq \hat{\mathbf{U}}_X^\top \hat{\mathbf{U}}_X$$

- (3) Le  $R^2$  indique la part de la variation de  $\mathbf{Y}$  dans l'échantillon qui est expliquée par  $\mathbf{X}$ . Cependant notre objectif n'est pas d'expliquer des variations dans l'échantillon mais celle de la population (dont est tiré l'échantillon). Il en résulte qu'un  $R^2$  élevé n'est pas nécessairement un indicateur d'un bon modèle de régression et un  $R^2$  faible n'est pas non plus un argument en défaveur du modèle considéré.
- (4) Il est toujours possible de trouver une matrice de régresseurs  $\mathbf{X}$  pour laquelle  $R^2 = 1$ , il suffit de prendre  $n$  vecteurs linéairement indépendants. En effet, un tel ensemble de vecteurs génère tout l'espace  $\mathbb{R}^n$  de sorte que tout vecteur  $\mathbf{Y} \in \mathbb{R}^n$  peut s'écrire comme une combinaison linéaire exacte des colonnes de  $\mathbf{X}$ .

6.  $R^2$  AJUSTÉ

Étant donné que le  $R^2$  augmente avec le nombre de régresseurs, une mesure alternative pour juger de la qualité de la régression est le  $R^2$  *ajusté*,

$$\begin{aligned}\bar{R}^2 &= 1 - \frac{n-1}{n-K}(1-R^2) \\ &= 1 - \frac{\hat{\mathbf{U}}^\top \hat{\mathbf{U}} / (n-K)}{\mathbf{Y}^\top \mathbf{M}_1 \mathbf{Y} / (n-1)}\end{aligned}$$

Le  $R^2$  ajusté diminue la qualité de ajustement lorsque le nombre de régresseurs augmente relativement au nombre d'observations de sorte que  $\bar{R}^2$  peut diminuer avec le nombre de régresseurs. Cependant il n'y a pas vraiment d'argument fort pour utiliser une telle mesure de l'ajustement.