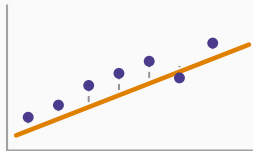


# Introduction à l'économétrie

## Chapitre 9 : Estimation et R carré ajusté



---

Jaouad Madkour

Faculté d'Économie et de Gestion de Tanger · 27 août 2026

Le défaut du R carré

Sur nos données

Ce qu'il faut retenir

## Le défaut du R carré

---

### Le $R^2$ ne peut jamais baisser

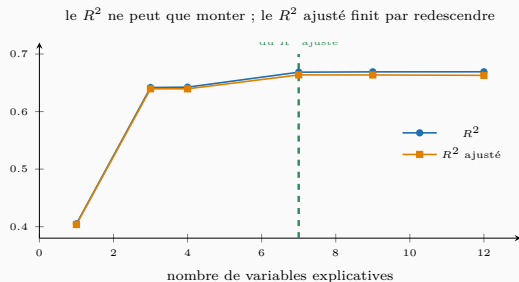
Ajouter une variable explicative ne peut pas augmenter la somme des carrés résiduels.

Au pire, son coefficient estimé sort nul et le modèle est inchangé. Le  $R^2$  est donc **croissant par construction** avec le nombre de variables.

### La conséquence est radicale

Choisir le modèle de plus grand  $R^2$  revient à choisir **le modèle le plus long**, toujours.

En ajoutant assez de variables sans rapport, on atteint  $R^2 = 1$  — et l'on n'a rien expliqué du tout. Le  $R^2$  est donc inutilisable pour comparer des modèles de tailles différentes.



## Pourquoi le $R^2$ monte toujours

Ajouter une variable ne peut pas augmenter la somme des carrés résiduels : au pire, son coefficient sort nul. Le  $R^2$  est donc **croissant par construction** — et par conséquent inutilisable pour choisir un modèle.

## La correction

$$\overline{R}^2 = 1 - (1 - R^2) \frac{n - 1}{n - k - 1}$$

Elle facture chaque variable ajoutée. Une variable n'améliore le  $R^2$  ajusté que si elle apporte plus que ce qu'elle coûte en degré de liberté.

**Le  $R^2$  ajusté peut être négatif.** Ce n'est pas une erreur de calcul : cela signale un modèle qui prédit moins bien que la simple moyenne.

### Le $R^2$ ajusté

$$\overline{R}^2 = 1 - (1 - R^2) \frac{n - 1}{n - k - 1}$$

Chaque variable ajoutée réduit  $n - k - 1$ , donc gonfle le second facteur : elle est **facturée**.

### Ce que la facture signifie

Une variable n'améliore le  $R^2$  ajusté que si elle apporte **plus qu'elle ne coûte en degré de liberté**.

On peut montrer que cela équivaut à ce que son  $|t|$  dépasse 1 — un seuil bien plus permissif que 1,96. Le  $R^2$  ajusté est donc un critère **indulgent**, et le chapitre 20 en présentera de plus sévères.

### Ce qui distingue les deux

|                      | $R^2$                | $R^2$ ajusté             |
|----------------------|----------------------|--------------------------|
| Ajout d'une variable | monte toujours       | monte ou baisse          |
| Valeurs possibles    | $[0 ; 1]$            | peut être <b>négatif</b> |
| Usage                | décrire l'ajustement | comparer des modèles     |

### Un $R^2$ ajusté négatif n'est pas une erreur de calcul

Il signale un modèle qui prédit **moins bien** que la **simple moyenne** de  $y$ .

Cela arrive avec beaucoup de variables sur peu d'observations. Le message est clair : il faut retirer des variables, ou en collecter davantage.

Sur nos données

---



### Ce que chaque ajout apporte

| Modèle        | $k$ | $R^2$ | $R^2$ ajusté |
|---------------|-----|-------|--------------|
| M1 ancienneté | 1   | 0,405 | 0,404        |
| M2 + diplôme  | 3   | 0,642 | 0,640        |
| M3 + sexe     | 4   | 0,642 | 0,640        |
| M4 + région   | 7   | 0,668 | 0,664        |

### La lecture, ligne par ligne

Le **diplôme** est décisif : le  $R^2$  passe de 0,405 à 0,642. Il expliquait à lui seul un quart de la dispersion, comme l'ANOVA du chapitre 18 de l'inférence l'avait établi.

Le **sexe** n'apporte rien :  $R^2$  inchangé,  $R^2$  ajusté qui baisse au troisième chiffre. La variable coûte plus qu'elle ne rapporte.

La **région** apporte encore deux points et demi. Le marché du travail marocain n'est pas homogène géographiquement.

### La lecture concrète du progrès

| Modèle | $\hat{\sigma}_\varepsilon$ |
|--------|----------------------------|
| M1     | 2 008 dh                   |
| M2     | 1 561 dh                   |
| M4     | 1 508 dh                   |

### Elle dit la même chose, en dirhams

Le modèle complet se trompe typiquement de mille cinq cents dirhams sur un salaire individuel, contre deux mille pour le modèle initial.

**C'est un progrès réel, et il reste modeste.** Deux tiers de la dispersion des salaires tiennent à des facteurs absents du fichier : poste, performance, négociation. Aucune régression ne fera mieux avec ces six colonnes.

Ce qu'il faut retenir

---

### Six points

1. Le  $R^2$  **monte toujours** quand on ajoute une variable : il ne peut pas servir à choisir.
2.  $\bar{R}^2 = 1 - (1 - R^2) \frac{n - 1}{n - k - 1}$  : chaque variable est **facturée**.
3. Une variable améliore le  $R^2$  ajusté dès que son  $|t|$  dépasse 1 — critère indulgent.
4. Le  $R^2$  ajusté peut être **négatif** : le modèle fait alors moins bien que la moyenne.
5. Sur nos données : le diplôme est décisif, le sexe n'apporte rien, la région compte un peu.
6.  $\hat{\sigma}_\epsilon$  passe de 2 008 à 1 508 dirhams — un progrès réel et modeste.

### La suite

Avec plusieurs variables, une question nouvelle se pose : tester chaque coefficient séparément suffit-il ?

Le chapitre 10 montre que non — et que trois tests différents cohabitent, dont les réponses peuvent diverger sans se contredire.