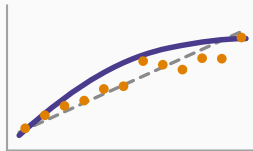


Économétrie avancée

Chapitre 2 : Le modèle de régression simple



Jaouad Madkour

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

Avant de commencer

Le modèle et son estimation

Les hypothèses de la régression simple

Qualité de l'ajustement

Formes fonctionnelles usuelles

Synthèse

Avant de commencer

Pourquoi ce chapitre ?

La régression simple explique une variable y par une seule variable x . Ses limites sont réelles (chapitre 3 les lèvera), mais elle introduit, sur le cas le plus simple possible, tout le vocabulaire et la logique d'estimation qui structureront l'ensemble du cours.

Le modèle et son estimation

Modèle de régression simple

$$y = \beta_0 + \beta_1 x + u,$$

où y est la variable **expliquée** (dépendante), x la variable **explicative** (indépendante), β_0, β_1 des paramètres inconnus de la population, et u le **terme d'erreur** — tous les autres facteurs, non observés, qui affectent y .

L'estimateur des moindres carrés ordinaires

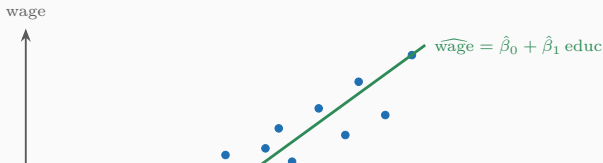
MCO (moindres carrés ordinaires)

À partir d'un échantillon $(x_i, y_i)_{i=1, \dots, n}$, on choisit $\hat{\beta}_0, \hat{\beta}_1$ qui **minimisent** la somme des carrés des résidus $\hat{u}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$:

$$\min_{\hat{\beta}_0, \hat{\beta}_1} \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2.$$

Les conditions du premier ordre (chapitre 3, cours d'analyse mathématique) donnent les formules explicites

$$\hat{\beta}_1 = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2}, \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}.$$



Les hypothèses de la régression simple

Quand les MCO sont-ils fiables ?

Hypothèses SLR

- SLR.1 : le modèle est linéaire en ses paramètres.
- SLR.2 : échantillonnage aléatoire dans la population.
- SLR.3 : x varie dans l'échantillon ($\sum_i (x_i - \bar{x})^2 > 0$) – sinon $\hat{\beta}_1$ n'est même pas calculable.
- SLR.4 : espérance conditionnelle nulle, $\mathbb{E}[u \mid x] = 0$ – en moyenne, les facteurs inobservés ne sont pas liés à x .

Non-biais des MCO

Sous SLR.1 à SLR.4, $\mathbb{E}[\hat{\beta}_0] = \beta_0$ et $\mathbb{E}[\hat{\beta}_1] = \beta_1$: en moyenne sur tous les échantillons possibles, l'estimateur MCO **visé juste**.

SLR.4 est l'hypothèse qui peut tout faire échouer

SLR.4 est rarement garantie en pratique. Si un facteur omis (l'aptitude naturelle, dans l'exemple du salaire) influence à la fois y et x , alors $\mathbb{E}[u \mid x] \neq 0$ et l'estimateur MCO est **biaisé** – il capture, en partie, l'effet du facteur omis. C'est précisément ce que la régression **multiple** (chapitre 3) cherche à corriger en contrôlant explicitement pour davantage de facteurs.

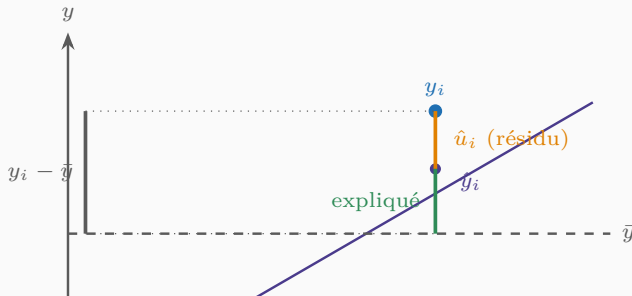
Qualité de l'ajustement

Décomposition de la variance et R^2

En notant $SST = \sum_i (y_i - \bar{y})^2$ (variation totale), $SSE = \sum_i (\hat{y}_i - \bar{y})^2$ (variation expliquée) et $SSR = \sum_i \hat{u}_i^2$ (variation résiduelle), on a $SST = SSE + SSR$, d'où

$$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST} \in [0, 1].$$

R^2 mesure la proportion de la variation de y expliquée par x .



Formes fonctionnelles usuelles

Le modèle log-linéaire

Interprétation en pourcentage

Avec $\log(y) = \beta_0 + \beta_1 x + u$, $\beta_1 \times 100$ s'interprète comme le **pourcentage** de variation de y pour une unité de plus de x (semi-élasticité) — utile quand l'effet de x sur y est mieux décrit en termes relatifs qu'en unités absolues.

Exemple (suite)

$\log(\text{salaire}) = 0,584 + 0,083 \text{ éducat}$, $R^2 = 0,186$: chaque année d'éducation supplémentaire est associée à une hausse d'environ 8,3% du salaire — le fameux « rendement de l'éducation ». Notons que R^2 (18,6%) porte ici sur $\log(\text{salaire})$, pas sur le salaire lui-même : les deux régressions ne sont pas directement comparables sur ce critère.

Application en finance

Le modèle log-linéaire est omniprésent en finance : les **rendements** eux-mêmes sont, par construction, des variations logarithmiques de prix ($r_t = \log(P_t/P_{t-1})$). Le modèle de marché du chapitre 1 (économétrie financière), $R_i = \alpha_i + \beta_i R_M + \varepsilon_i$, **est** une régression simple — et β_i s'interprète directement comme l'élasticité du rendement de l'actif au rendement du marché.

Synthèse

L'essentiel du chapitre 2

- MCO : minimise $\sum \hat{u}_i^2$, formules explicites $\hat{\beta}_1 = S_{xy}/S_{xx}$, $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$.
- Non-biais garanti sous SLR.1-SLR.4; SLR.4 (exogénéité de x) est l'hypothèse la plus fragile en pratique.
- $R^2 = SSE/SST$ mesure l'ajustement, mais ne doit pas être surinterprété.
- Le modèle log-linéaire donne une interprétation en pourcentage — la forme naturelle en finance, où l'on travaille en rendements.

Vers le chapitre 3

Contrôler pour **plusieurs** facteurs à la fois, au lieu d'un seul, est le moyen le plus direct de rapprocher SLR.4 de la réalité. C'est l'objet de la régression multiple.