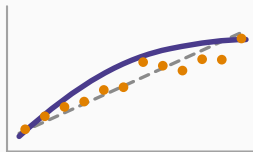


Économétrie avancée

Chapitre 3 : Régression multiple — estimation



Jaouad Madkour

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

Avant de commencer

Le modèle et son estimation

Les hypothèses de Gauss-Markov

Qualité de l'ajustement

Synthèse

Avant de commencer

Pourquoi ce chapitre ?

La régression simple ne peut contrôler qu'un seul facteur à la fois — une hypothèse SLR.4 rarement crédible. La régression **multiple** contrôle simultanément plusieurs facteurs, rapprochant l'analyse *ceteris paribus* d'une véritable comparaison « toutes choses égales par ailleurs ».

Le modèle et son estimation

Spécification et écriture matricielle

Modèle de régression multiple

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u,$$

ou, sous forme matricielle (chapitre 1, algèbre linéaire), $y = X\beta + u$ où X est la matrice $n \times (k + 1)$ des régresseurs (une colonne de 1 pour l'intercept, puis une colonne par x_j).

L'estimateur MCO en notation matricielle

En minimisant $\sum_i \hat{u}_i^2 = (y - X\hat{\beta})^\top (y - X\hat{\beta})$, les conditions du premier ordre donnent

$$\hat{\beta} = (X^\top X)^{-1} X^\top y,$$

qui n'existe que si $X^\top X$ est **inversible** (chapitre 4, algèbre linéaire) — c'est-à-dire si aucun régresseur n'est combinaison linéaire exacte des autres.

Chaque β_j , toutes choses égales par ailleurs

$\hat{\beta}_j$ mesure l'effet d'une hausse d'une unité de x_j sur y , **en maintenant fixes tous les autres régresseurs du modèle** — l'interprétation *ceteris paribus* que la régression simple ne pouvait offrir qu'imparfaitement.

Les hypothèses de Gauss-Markov

Hypothèses MLR

- **MLR.1–2** : linéarité en les paramètres, échantillonnage aléatoire.
- **MLR.3** (absence de colinéarité parfaite) : aucun x_j n'est une combinaison linéaire exacte des autres régresseurs.
- **MLR.4** (exogénéité) : $\mathbb{E}[u \mid x_1, \dots, x_k] = 0$.
- **MLR.5** (homoscédasticité) : $\text{Var}(u \mid x_1, \dots, x_k) = \sigma^2$, constante.

Théorème de Gauss-Markov

Sous MLR.1 à MLR.4, l'estimateur MCO est **non biaisé** : $\mathbb{E}[\hat{\beta}_j] = \beta_j$. Sous MLR.1 à MLR.5 (l'ensemble complet), l'estimateur MCO est en outre le meilleur estimateur linéaire non biaisé (**BLUE** — *Best Linear Unbiased Estimator*) : parmi tous les estimateurs linéaires et non biaisés de β_j , il a la **variance la plus faible**.

Que se passe-t-il si l'on oublie un facteur pertinent ?

Si le « vrai » modèle contient x_1 et x_2 , mais qu'on estime une régression n'incluant que x_1 , l'estimateur de son coefficient est biaisé dès que x_1 et x_2 sont corrélés — l'estimateur simple capture alors, en partie, l'effet de x_2 « transmis » via sa corrélation avec x_1 . C'est le **biais de variable omise**, la généralisation directe du problème rencontré au chapitre 2.

Corrélation entre régresseurs : pas de biais, mais de la variance

Une forte corrélation entre régresseurs (**multicollinéarité**) ne biaise **pas** les MCO (MLR.4 reste satisfaite), mais **gonfle la variance** des estimateurs : il devient statistiquement difficile de démêler l'effet propre de chaque variable corrélée. Un cas extrême — une corrélation parfaite — viole directement MLR.3 et rend l'estimation impossible (chapitre 3, algèbre linéaire : le rang de X chute, $X^\top X$ n'est plus inversible).

Qualité de l'ajustement

R^2 ajusté

Le R^2 **augmente mécaniquement** avec chaque variable ajoutée, utile ou non. Le R^2 ajusté pénalise l'ajout de régresseurs :

$$\bar{R}^2 = 1 - \frac{SSR/(n - k - 1)}{SST/(n - 1)},$$

et peut **diminuer** si la variable ajoutée contribue trop peu à réduire SSR – un critère de comparaison de modèles plus fiable que le R^2 simple.

Application en finance : le modèle multifactoriel

Le modèle multifactoriel du chapitre 1 (économétrie financière), $R_i = \alpha_i + \sum_k \beta_{ik} F_k + \varepsilon_i$, est une régression multiple standard : chaque β_{ik} s'interprète *ceteris paribus* – la sensibilité de l'actif au facteur k , en tenant les autres facteurs constants. Le biais de variable omise y prend un sens concret : oublier un facteur de risque pertinent (par exemple un facteur sectoriel) dans le modèle fausse l'estimation des sensibilités aux facteurs retenus, dès que ce facteur omis est corrélé avec eux.

Synthèse

L'essentiel du chapitre 3

- $\hat{\beta} = (X^T X)^{-1} X^T y$; chaque coefficient s'interprète *ceteris paribus*.
- Gauss-Markov : non-biais sous MLR.1-4; BLUE (variance minimale parmi les estimateurs linéaires non biaisés) sous MLR.1-5.
- Le biais de variable omise apparaît dès qu'un facteur pertinent, corrélé aux régresseurs inclus, est oublié.
- La multicollinéarité gonfle la variance sans biaiser; le R^2 ajusté pénalise l'ajout de régresseurs peu utiles.

Vers le chapitre 4

Estimer un coefficient ne suffit pas : encore faut-il savoir s'il est **statistiquement significatif**, et construire des intervalles de confiance fiables. C'est l'objet de l'inférence.