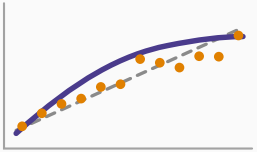
Économétrie avancée

Annexe E : Le modèle de régression linéaire sous forme matricielle



Jaouad Madkour

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

Plan du chapitre

Avant de commencer

Pourquoi cette annexe

Pourquoi ce chapitre ?

Ce chapitre referme le cours en reformulant, en notation matricielle compacte, tout ce qui a été établi progressivement depuis le chapitre 2 : l'estimateur MCO, son non-biais, sa matrice de variance-covariance, le théorème de Gauss-Markov, et l'inférence exacte sous normalité. Rien de nouveau sur le plan économique — mais une vision d'ensemble, unifiée, du cours entier.

Le modèle et l'estimateur MCO

Écriture matricielle

Le modèle sous forme matricielle

Avec $\mathbf{y}$ $(n \times 1)$ le vecteur des observations, $\mathbf{X}$ $(n \times k)$ la matrice des régresseurs (première colonne de 1 pour l'intercept) et $\mathbf{u}$ $(n \times 1)$ le vecteur des erreurs, le modèle linéaire s'écrit en une seule équation :

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{u}.$$

Chacune des n équations individuelles $y_t = \mathbf{x}_t\beta + u_t$ (chapitre 3) est une ligne de cette écriture compacte.

L'estimateur MCO en une formule

Minimiser $SSR(\mathbf{b}) = (\mathbf{y} - \mathbf{X}\mathbf{b})'(\mathbf{y} - \mathbf{X}\mathbf{b})$ conduit aux équations normales $\mathbf{X}'\mathbf{X}\widehat{\beta} = \mathbf{X}'\mathbf{y}$ — l'équivalent matriciel exact des conditions de premier ordre du chapitre 3 — dont la solution, quand $\mathbf{X}'\mathbf{X}$ est inversible, est

$$\widehat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}.$$

Cette formule unique condense tout le système des k équations à k inconnues du chapitre 3.

Le rôle exact du rang plein

$\mathbf{X}'\mathbf{X}$ est inversible si et seulement si $\mathrm{rang}(\mathbf{X}) = k$ — la traduction matricielle directe de MLR.3 (absence de colinéarité parfaite, chapitre 2). Sans cette condition, $\widehat{\beta}$ n'est simplement pas défini : ce n'est pas un problème d'inférence, mais d'existence même de l'estimateur.

Valeurs ajustées, résidus, et orthogonalité

$\hat{\mathbf{y}} = \mathbf{X}\widehat{\beta}$, $\hat{\mathbf{u}} = \mathbf{y} - \hat{\mathbf{y}}$, et la condition de premier ordre équivaut à $\mathbf{X}'\hat{\mathbf{u}} = \mathbf{0}$: puisque la première colonne de $\mathbf{X}$ est un vecteur de 1, ceci implique immédiatement que les résidus somment à zéro et que chaque régresseur est en covariance d'échantillon nulle avec les résidus — deux propriétés du chapitre 3, ici obtenues en une ligne.

Propriétés à échantillon fini

Hypothèses E.1–E.3 et non-biais

Sous linéarité (E.1), exogénéité stricte $E(\mathbf{u} \mid \mathbf{X}) = \mathbf{0}$ (E.2, généralisation matricielle de MLR.3-4) et rang plein (E.3), on écrit

$$\widehat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u},$$

d'où $E(\widehat{\beta} \mid \mathbf{X}) = \beta$ immédiatement, puisque $E(\mathbf{u} \mid \mathbf{X}) = \mathbf{0}$. Toute la démonstration du non-biais MCO (chapitre 3) tient en cette seule ligne de calcul matriciel.

Homoscédasticité et absence d'autocorrélation réunies

Sous E.4, $\mathrm{Var}(\mathbf{u} \mid \mathbf{X}) = \sigma^2 \mathbf{I}_n$ (chapitre 13), on obtient

$$\mathrm{Var}(\hat{\beta} \mid \mathbf{X}) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}.$$

La variance de $\hat{\beta}_j$ est σ^2 fois le j-ième terme diagonal de $(\mathbf{X}'\mathbf{X})^{-1}$; les covariances entre estimateurs, nécessaires pour tester des combinaisons linéaires de coefficients (chapitre 4), sont directement les termes hors diagonale — plus besoin de la formule ad hoc du chapitre 3 pour chaque cas particulier.

MCO = BLUE, démonstration matricielle

Pour tout estimateur linéaire non biaisé $\tilde{\beta} = \mathbf{A}\mathbf{y}$ (avec $\mathbf{A}\mathbf{X} = \mathbf{I}_k$, condition de non-biais), on montre que

$$\mathrm{Var}(\tilde{\beta} \mid \mathbf{X}) - \mathrm{Var}(\hat{\beta} \mid \mathbf{X}) = \sigma^2 \mathbf{A}'\mathbf{M}\mathbf{A},$$

où $\mathbf{M} = \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ est **idempotente** (chapitre 13) — donc $\mathbf{A}'\mathbf{M}\mathbf{A}$ est semi-définie positive. Pour toute combinaison linéaire $\mathbf{c}'\beta$, $\mathrm{Var}(\mathbf{c}'\tilde{\beta}) \geq \mathrm{Var}(\mathbf{c}'\hat{\beta})$: les MCO sont **BLUE**, dans toute leur généralité, pas seulement coefficient par coefficient.

La matrice $\mathbf{M}$, déjà rencontrée

$\mathbf{M}$ est exactement la matrice idempotente de rang $n - k$ introduite au chapitre 13 comme préfiguration des résidus : $\hat{\mathbf{u}} = \mathbf{M}\mathbf{y} = \mathbf{M}\mathbf{u}$. Elle joue ici un double rôle — produire les résidus, et démontrer l'optimalité des MCO — ce qui illustre pourquoi les outils du chapitre 13 étaient nécessaires avant d'aborder ce résultat.

Estimation de σ^2

Non-biais de $\hat{\sigma}^2$

Avec $\hat{\sigma}^2 = \hat{\mathbf{u}}'\hat{\mathbf{u}}/(n-k)$, en utilisant $\hat{\mathbf{u}} = \mathbf{M}\mathbf{u}$, $\mathrm{tr}(\mathbf{M}) = n-k$ (chapitre 13) et la propriété $E[\mathrm{tr}(\mathbf{M}\mathbf{u}\mathbf{u}') \mid \mathbf{X}] = \sigma^2 \,\mathrm{tr}(\mathbf{M})$, on obtient $E(\hat{\sigma}^2 \mid \mathbf{X}) = \sigma^2$ — la démonstration générale, en quelques lignes, de ce qui était admis sans preuve complète au chapitre 3.

Inférence exacte sous normalité

Normalité de $\widehat{\beta}$ et lois exactes

Le modèle linéaire classique en notation matricielle

Ajoutant $\mathbf{u} \mid \mathbf{X} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ (E.5, hypothèse MLR.6 du chapitre 4), on obtient

$$\widehat{\beta} \mid \mathbf{X} \;\sim\; \mathcal{N}\big(\beta,\; \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}\big),$$

puisque $\widehat{\beta}$ est une transformation linéaire de $\mathbf{u}$ (propriété de la loi normale multivariée, chapitre 13).

D'où viennent exactement les lois t et F

Trois ingrédients, tous établis au chapitre 13, se combinent : (i) $(\widehat{\beta}_j - \beta_j)/sd(\widehat{\beta}_j) \sim \mathcal{N}(0,1)$; (ii) $(n-k)\widehat{\sigma}^2/\sigma^2 = (\mathbf{u}/\sigma)'\mathbf{M}(\mathbf{u}/\sigma) \sim \chi^2_{n-k}$ car $\mathbf{M}$ est idempotente de rang $n-k$; (iii) $\widehat{\beta}$ et $\widehat{\sigma}^2$ sont **indépendants**, car $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{M} = \mathbf{0}$. En combinant (i), (ii) et (iii) selon la définition même de la loi de Student (chapitre 13),

$$\frac{\widehat{\beta}_j - \beta_j}{\mathrm{se}(\widehat{\beta}_j)} \;\sim\; t_{n-k}.$$

Le test t du chapitre 4, enfin complètement démontré

Cette dérivation referme une boucle ouverte depuis le chapitre 4 : le résultat

Efficacité et lien avec le maximum de vraisemblance

MCO = estimateur à variance minimale, MCO = MV

Sous les hypothèses E.1–E.5, la borne de Cramér-Rao pour tout estimateur non biaisé de β est exactement $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$: les MCO sont donc **optimaux parmi tous les estimateurs non biaisés**, pas seulement parmi les estimateurs linéaires (extension du théorème de Gauss-Markov). De plus, sous normalité, maximiser la vraisemblance revient exactement à minimiser $SSR(\mathbf{b})$: MCO et maximum de vraisemblance **coïncident** — sauf pour σ^2, où l'estimateur MV (SSR/n) est biaisé, contrairement à $\hat{\sigma}^2 = SSR/(n-k)$ systématiquement préféré en pratique.

Synthèse générale du cours

Ce qu'il faut retenir

L'essentiel du chapitre 14

- $\widehat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ condense en une formule tout le système d'équations normales du chapitre 3 ; son existence exige le rang plein de $\mathbf{X}$.
- Non-biais, matrice de variance-covariance $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$, et théorème de Gauss-Markov (BLUE) se démontrent en quelques lignes à partir des outils du chapitre 13 — idempotence, positivité semi-définie, propriétés de l'espérance et de la variance matricielles.
- Sous normalité, les lois t et F exactes du chapitre 4 découlent directement des propriétés des matrices idempotentes appliquées au vecteur d'erreurs.
- Les MCO sont l'estimateur à variance minimale parmi **tous** les estimateurs non biaisés (pas seulement linéaires) sous normalité, et coïncident avec le maximum de vraisemblance pour β.

Fin de ce second bloc du cours

Les chapitres 6 à 14 ont approfondi et généralisé les fondations posées aux chapitres 1 à 5 : formes fonctionnelles avancées, variables qualitatives, hétéroscédasticité, problèmes de spécification, séries temporelles, et enfin l'appareil probabiliste et matriciel complet qui sous-tend l'ensemble. Les chapitres suivants (variables instrumentales, données de panel) prolongeront ce cadre vers les situations où l'exogénéité stricte elle-même doit être abandonnée.