ÉCONOMÉTRIE 2
UGA, M1 MIASH-BDA, S2

RÉGRESSION LINÉAIRE
ESTIMATEUR DES MOINDRES CARRÉS

(

RÉVISION DES PROPRIÉTÉS ET MÉTHODES D'INFÉRENCE ASYMPTOTIQUES)

(CETTE VERSION : 28 FÉVRIER 2023)

MICHAL URDANIVIA [1]

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

TABLE DES MATIÈRES

1. Le modèle

On s'intéresse à la relation entre une variable $Y \in \mathbb{R}$, appelée *variable dépendante*, et un vecteur $X \in \mathbb{R}^K$, de variables appelées régresseurs. Pour cela nous disposons de données $\{(Y_i, X_i)\}_{i=1}^n$, et le modèle que nous considérons est un modèle de régression linéaire défini par les hypothèses suivantes.

Condition C1. Les données $\{(Y_i, X_i), i = 1, ..., n\}$ sont un échantillon i.i.d.

Condition C2. Y_i et X_i vérifient,

$$Y_i = X_i^\top \beta + U_i \quad i = 1, ..., n$$

où U_i est une variable inobservée(ou terme d'erreur) vérifiant $\mathrm{E}(U_i) = 0$.

Condition C3. X_i est (faiblement)exogène [2] par rapport à U_i,

$$\mathrm{E}(X_i U_i) = 0$$

Condition C4. La matrice $\mathrm{E}(X_i X_i^\top)$ est finie et définie positive.

Condition C5. $\mathrm{E}(X_{i,k}^4) < \infty$, pour tout $k = 1, ..., K$.

Condition C6. $\mathrm{E}(U_i^4) < \infty$

Condition C7. $\mathrm{E}(U_i^2 X_i X_i^\top)$ est définie positive.

2. Convergence

Rappelons qu'une écriture de l' estimateur des moindres carrés est,

$$\hat{\beta}_n = \left(\sum_{i=1}^n X_i X_i^\top \right)^{-1} \sum_{i=1}^n X_i Y_i$$

Cet estimateur est convergent pour β si $\hat{\beta}_n \xrightarrow{p} \beta$ quand $n \to \infty$, ce qui est établi par la propriété suivant.

Propriété 2.1. ***(Convergence de l'estimateur des moindres carrés)*** *Sous les hypothèses C1 - C4, $\hat{\beta}_n \xrightarrow{p} \beta$ quand $n \to \infty$.*

Démonstration. $\hat{\beta}_n$ peut s'écrire,

$$\hat{\beta}_n = \beta + \left(n^{-1} \sum_{i=1}^n X_i X_i^\top \right)^{-1} n^{-1} \sum_{i=1}^n X_i U_i \tag{2.1}$$

2. On pourrait aussi supposer une *exogénéité forte* avec $\mathrm{E}(U_i|X_i) = 0$ auquel cas β dans $Y_i = X_i^\top \beta + U_i$ seraient les paramètres d'un modèle de régression linéaire, à savoir d'un modèle tel que $\mathrm{E}(Y_i|X_i) = X_i^\top \beta$.

Les termes U_i's et les termes X_iU_i's sont i.i.d. sous l'hypothèse C1. Dans ce cas, par la loi faible de grands nombres [3],

$$n^{-1}\sum_{i=1}^{n} X_iU_i \xrightarrow{p} \mathrm{E}(X_iU_i) = 0$$

Où l'on utilise C3. Dans la mesure où $E(X_iX_i^\top)$ est finie sous l'hypothèse C4 nous avons par la loi faible des grand nombres,

$$n^{-1}\sum_{i=1}^{n} X_iX_i^\top \xrightarrow{p} \mathrm{E}(X_iX_i^\top)$$

et comme $E(X_iX_i^\top)$ est définie positive, nous avons par le théorème de Slutsky,

$$\left(n^{-1}\sum_{i=1}^{n} X_iX_i^\top\right)^{-1} \xrightarrow{p} \left(\mathrm{E}(X_iX_i^\top)\right)^{-1} \tag{2.2}$$

Et par conséquent,

$$\left(n^{-1}\sum_{i=1}^{n} X_iX_i^\top\right)^{-1} n^{-1}\sum_{i=1}^{n} X_iU_i \xrightarrow{p} 0$$

et donc,

$$\hat{\beta}_n \xrightarrow{p} \beta$$

□

3. Soit X une variable aléatoire, et définissons,

$$X^+ = \max(0, X)$$
$$X^- = \max(0, -X),$$

de sorte que,

$$X = X^+ - X^-.$$

On note que X^+ et X^- sont toutes les deux des variables aléatoires positives. Quand au moins une des conditions suivantes est satisfaite :

(i) $\mathrm{E}(X^+) < \infty$, ou,

(ii) $\mathrm{E}(X^-) < \infty$,

la valeur espérée de X est donnée par,

$$\mathrm{E}(X) = \mathrm{E}(X^+) - \mathrm{E}(X^-).$$

$\mathrm{E}(X)$ n'est pas définie quand $\mathrm{E}(X^+) = \infty$ et $\mathrm{E}(X^-) = \infty$(ainsi nous excluons/interdisons $\infty - \infty$. Comme,

$$|X| = X^+ + X^-,$$

nous avons que $\mathrm{E}(|X|) < \infty$ si et seulement si $\mathrm{E}(X^+) < \infty$ et $\mathrm{E}(X^-) < \infty$. Quand nous disons que $\mathrm{E}(X) = \mu$ pour un certain μ, nous supposons donc que soit $\mathrm{E}(X^+) < \infty$ ou $\mathrm{E}(X^-) < \infty$ pour que $\mathrm{E}(X)$ soit définie. Si μ est fini, il doit être vrai que $\mathrm{E}(X^+) < \infty$ et $\mathrm{E}(X^-) < \infty$, et par conséquent que $\mathrm{E}(|X|) < \infty$.

3. DISTRIBUTION ASYMPTOTIQUE

Le résultat suivant établit la distribution asymptotique de l'estimateur des moindres carrés.

Propriété 3.1. ***(Normalité asymptotique)*** *Sous les hypothèses C1-C7,*

$$n^{1/2}(\hat{\beta}_n - \beta) \xrightarrow{d} \mathcal{N}(0, \mathrm{V})$$

où

$$\mathrm{V} = Q^{-1}\Omega Q^{-1}, \quad Q = \mathrm{E}(X_i X_i^\top), \quad \Omega = \mathrm{E}(U_i^2 X_i X_i^\top)$$

Démonstration. Nous avons en utilisant (2.1),

$$n^{1/2}(\hat{\beta}_n - \beta) = \left(n^{-1}\sum_{i=1}^{n} X_i X_i^\top\right)^{-1} n^{-1/2}\sum_{i=1}^{n} X_i U_i$$

Commençons en considérant le terme $n^{-1/2}\sum_{i=1}^{n} X_i U_i$. Par C3, $\mathrm{E}(X_i U_i) = 0$. Ensuite, considérons $\mathrm{Var}(X_i U_i) = \mathrm{E}(U_i^2 X_i X_i^\top)$. L'élément (r, s), $r, s = 1, ..., K$, de $\mathrm{Var}(X_i U_i) = \mathrm{E}(U_i^2 X_i X_i^\top)$ est $\mathrm{E}(U_i^2 X_{i,r} X_{i,s})$. Par l'inégalité de Cauchy-Schwartz et sous les hypothèses C5 , C6,

$$\mathrm{E}\left(\left|U_i^2 X_{i,r} X_{i,s}\right|\right) \leq \left[\mathrm{E}(U_i^4)\,\mathrm{E}(X_{i,r}^2 X_{i,s}^2)\right]^{1/2} \leq \left[\mathrm{E}(U_i^4)^{1/2}\,\mathrm{E}(X_{i,r}^4)\,\mathrm{E}(X_{i,s}^4))\right]^{1/4} < \infty$$

Par le théorème central-limite,

$$n^{-1/2}\sum_{i=1}^{n} X_i U_i \xrightarrow{d} \mathcal{N}\left(0, \mathrm{E}(U_i^2 X_i X_i^\top)\right) = \mathcal{N}(0, \Omega) \tag{3.1}$$

Finalement (2.2), (3.1) et le théorème de convergence de Cramer(son extension multivariée) impliquent que,

$$\left(n^{-1}\sum_{i=1}^{n} X_i X_i^\top\right)^{-1} n^{-1/2}\sum_{i=1}^{n} X_i U_i \xrightarrow{d} Q^{-1}\mathcal{N}(0, \Omega) = \mathcal{N}(0, Q^{-1}\Omega Q^{-1})$$

□

Remarque 1. Les hypothèses de la propriété 3.1 n'excluent pas le cas où la variance conditionnelle des U_i's est une fonction de X_i, i.e. il est possible que les termes d'erreur U_i's soient *hétéroscédastiques* : $\mathrm{E}(U_i^2|X_i) = \sigma^2(X_i)$ pour une fonction $\sigma^2 : \mathbb{R}^K \mapsto \mathbb{R}$.

Remarque 2. La matrice de variances-covariances asymptotique de $\hat{\beta}_n$ est donnée par la formule "en sandwich",

$$\mathrm{V} = \left(\mathrm{E}(X_i X_i^\top)\right)^{-1} \mathrm{E}(U_i^2 X_i X_i^\top) \left(\mathrm{E}(X_i X_i^\top)\right)^{-1}$$

Si nous imposons la condition que $\mathrm{E}(U_i^2|X_i) = \sigma^2$, alors V se simplifie en la matrice des variances-covariances *homoscédastique*,

$$\mathrm{V} = \sigma^2\left(\mathrm{E}(X_i X_i^\top)\right)^{-1} \tag{3.2}$$

En effet par la règle des espérances itérées,

$$\mathrm{E}(U_i^2 X_i X_i^\top) = \mathrm{E}\left(\mathrm{E}(U_i^2 X_i X_i^\top)|X_i)\right) = \mathrm{E}\left(X_i X_i^\top\, \mathrm{E}(U_i^2|X_i)\right) = \sigma^2\, \mathrm{E}(X_i X_i^\top)$$

ainsi dans ce cas,

$$\Omega = \sigma^2 Q \;\text{ et, }\; \mathrm{V} = Q^{-1}\Omega Q^{-1} = \sigma^2 Q^{-1} = \sigma^2\left(\mathrm{E}(X_i X_i^\top)\right)^{-1}$$

4. Estimation de la matrice des variances-covariances

A partir d'un estimateur de β, nous pouvons construire les résidus $\hat{U}_i = Y_i - X_i^\top \hat{\beta}_n$. Considérons l'estimateur suivant de V obtenu par application du principe d'analogie,

$$\hat{\mathrm{V}}_n = \hat{Q}_n^{-1}\hat{\Omega}_n \hat{Q}_n^{-1}$$

où,

$$\hat{Q}_n = n^{-1}\sum_{i=1}^n X_i X_i^\top \;,\; \hat{\Omega}_n = n^{-1}\sum_{i=1}^n \hat{U}_i^2 X_i X_i^\top$$

Nous avons déjà montré que $\hat{Q}_n^{-1} \xrightarrow{d} Q^{-1}$(c.f., (2.2)). Considérons maintenant $\hat{\Omega}_n$. Nous pouvons écrire ici,

$$\hat{U}_i = U_i - X_i(\hat{\beta}_n - \beta)$$

Il en résulte que,

$$n^{-1}\sum_{i=1}^n \hat{U}_i^2 X_i X_i^\top = n^{-1}\sum_{i=1}^n U_i^2 X_i X_i^\top - 2R_{1,n} + R_{2,n} \tag{4.1}$$

où,

$$R_{1,n} = n^{-1}\sum_{i=1}^n \left((\hat{\beta}_n - \beta) X_i U_i\right) X_i X_i^\top \;,\; R_{2,n} = n^{-1}\sum_{i=1}^n \left((\hat{\beta}_n - \beta) X_i\right)^2 X_i X_i^\top$$

Sous les hypothèses de la propriété 3.1, $\mathrm{E}(U_i^2 X_i X_i^\top)$ est finie, comme cela a été montré dans la démonstration de la propriété. Par conséquent, par la loi faible des grand nombres,

$$n^{-1}\sum_{i=1}^n U_i^2 X_i X_i^\top \xrightarrow{p} \mathrm{E}(U^2 X_i X_i^\top)$$

En outre, il est possible de montrer que $R_{1,n}$ et $R_{2,n}$ convergent en probabilité vers zéro(c.f., annexe) de sorte que,

$$\hat{\mathrm{V}}_n \xrightarrow{p} \mathrm{V}$$

L'estimateur de la matrice des variances-covariances $\hat{\mathrm{V}}_n = \hat{Q}_n^{-1}\hat{\Omega}_n\hat{Q}_n^{-1}$, qui est ainsi donné par une formule "en sandwich" est un estimateur convergent que les termes d'erreur soient homoscédastiques ou hétéroscédastiques. Il est fréquent de l'appeler *estimateur convergent robuste à l'hétéroscédasticité*, ou *estimateur robuste de White*(car il fut suggéré par (**?**))

5. Intervalles de confiance asymptotiques

Dans cette section nous nous intéressons aux intervalles de confiance pour les éléments de β. Considérons l'intervalle de confiance suivant pour β_k, $k = 1, ..., K$,

$$\mathrm{CI}_{n,k,1-\alpha} = \left[\hat{\beta}_{n,k} - z_{1-\alpha/2}\sqrt{\left[\hat{\mathrm{V}}_n\right]_{k,k}/n}, \hat{\beta}_{n,k} + z_{1-\alpha/2}\sqrt{\left[\hat{\mathrm{V}}_n\right]_{k,k}/n}\right]$$

où $z_{1-\alpha/2}$ est le quantile $1-\alpha/2$ de la distribution normale standard et $\left[\hat{\mathrm{V}}_n\right]_{k,k}$ est l'élément (k,k) de la matrice $\hat{\mathrm{V}}_n$. Nous allons montrer que $\mathrm{P}\left(\beta_k \in \mathrm{CI}_{n,k,1-\alpha}\right) \to 1-\alpha$ lorsque $n \to \infty$. Comme $n^{1/2}(\hat{\beta}_n - \beta) \xrightarrow{d} \mathcal{N}(0, V)$, et $\hat{\mathrm{V}}_n \xrightarrow{p} V$, il résulte des théorèmes de convergence Slutsky et de Cramer que,

$$\hat{\mathrm{V}}_n^{-1/2} n^{1/2}(\hat{\beta}_n - \beta) \xrightarrow{d} \mathrm{V}^{-1/2}\,\mathcal{N}(0,\mathrm{V}) = \mathcal{N}(0, \mathbf{I}_K)$$

et par conséquent,

$$\frac{\sqrt{n}(\hat{\beta}_{n,k} - \beta_k)}{\sqrt{\left[\hat{\mathrm{V}}_n\right]_{k,k}}} \xrightarrow{d} \mathcal{N}(0,1)$$

ce qui peut aussi s'écrire comme,

$$\mathrm{P}\left(\frac{\sqrt{n}(\hat{\beta}_{n,k} - \beta_k)}{\sqrt{\left[\hat{\mathrm{V}}_n\right]_{k,k}}} \leq z\right) \to \mathrm{P}(Z \leq z) \;\text{ pour tout }\; z \in \mathbb{R},$$

où Z est une variable aléatoire et $Z \sim \mathcal{N}(0,1)$. A présent,

$$\mathrm{P}(\beta_k \in \mathrm{CI}_{n,k,1-\alpha}) = \mathrm{P}\left(\frac{\sqrt{n}(\hat{\beta}_{n,k} - \beta_k)}{\sqrt{\left[\hat{\mathrm{V}}_n\right]_{k,k}}} \leq z_{1-\alpha/2}\right) \to \mathrm{P}(|Z| \leq z_{1-\alpha/2}) = 1-\alpha$$

Considérons, par exemple, le cas avec des termes d'erreur homoscédastiques. Nous avons vu que dans ce cas $\sqrt{n}(\hat{\beta}_n - \beta) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2\left(\mathrm{E}(XX^\top)\right)^{-1}\right)$. Comme $s^2 \xrightarrow{p} \sigma^2$, la matrice des variances-covariances peut être estimée par $s^2\left(\sum_{i=1}^n X_i X_i^\top\right)^{-1}$. Et l'intervalle de confiance pour β_k est alors,

$$\left[\hat{\beta}_{n,k} \pm z_{1-\alpha/2}\sqrt{\left[s^2\left(n^{-1}\sum_{i=1}^n X_i X_i^\top\right)^{-1}\right]_{k,k}/n}\right] = \left[\hat{\beta}_{n,k} \pm z_{1-\alpha/2}\sqrt{\left[s^2\left(\mathbf{X}^\top\mathbf{X}\right)\right]_{k,k}}\right]$$

qui est le même intervalle de confiance que celui à distance finie, sauf qu'on utilise ici les quantiles de la distribution normale standard plutôt que ceux de la loi de student.

6. Tests d'hypothèses

Dans cette section nous considérons les tests asymptotiques de l'hypothèse $H_0 : h(\beta) = 0$ contre l'alternative $H_1 : h(\beta) \neq 0$, où $h : \mathbb{R}^K \mapsto \mathbb{R}^q$ est une fonction continument dérivable dans un voisinage de β. La contrainte sous H_0 inclut le cas des contraintes linéaires de la forme $h(\beta) = \mathbf{R}\beta - r$, où $\mathbf{R}$ est une matrice $q \times K$ et r est un vecteur de taille q.
Considérons la *statistique de test de Wald*,

$$W_n = nh(\hat{\beta}_n)^\top \left(\widehat{\text{AsyVar}}\left(h(\hat{\beta}_n)\right)\right)^{-1} h(\hat{\beta}_n) = nh(\hat{\beta}_n)^\top \left(\frac{\delta h}{\delta \beta^\top}(\hat{\beta}_n)\hat{\text{V}}_n \frac{\delta h}{\delta \beta}(\hat{\beta}_n)^\top\right)^{-1} h(\hat{\beta}_n)$$

où AsyVar désigne la variance asymptotique. Le test asymptotique de taille α de $H_0 : h(\beta) = 0$ est alors défini par la règle,

$$\text{Rejeter } H_0 \text{ si } W_n > \chi^2_{q,1-\alpha}$$

où $\chi^2_{q,1-\alpha}$ est le quantile $(1-\alpha)$ de la distribution du χ^2_q. Un test s'appuyant sur W_n est dit convergent si $\text{P}(W_n > \chi^2_{q,1-\alpha} | H_1) \to 1$.

Propriété 6.1. *Sous les hypothèses C1-C6,*

(1) $\text{P}(W_n > \chi^2_{q,1-\alpha} | H_0) \to \alpha$.

(2) $\text{P}(W_n > \chi^2_{q,1-\alpha} | H_1) \to 1$.

Démonstration. (1) Comme $n^{1/2}(\hat{\beta}_n - \beta) \xrightarrow{d} \mathcal{N}(0, \text{V})$ et que $h(.)$ est continue en β, sous H_0, et en appliquant la méthode delta,

$$n^{1/2} h(\hat{\beta}_n) \xrightarrow{d} \mathcal{N}\left(0, \frac{\delta h}{\delta \beta^\top}(\beta)\, \text{V}\, \frac{\delta h}{\delta \beta}(\beta)^\top\right)$$

En outre,

$$\frac{\delta h}{\delta \beta^\top}(\hat{\beta}_n) \xrightarrow{p} \frac{\delta h}{\delta \beta^\top}(\beta) \text{ et, } \hat{\text{V}}_n \xrightarrow{p} \text{V}$$

Par la propriété de convergence de Cramer, sous H_0,

$$\left(\frac{\delta h}{\delta \beta^\top}(\hat{\beta}_n)\hat{\text{V}}_n \frac{\delta h}{\delta \beta}(\hat{\beta}_n)^\top\right)^{-1/2} n^{1/2} h(\hat{\beta}_n) \xrightarrow{d} \left(\frac{\delta h}{\delta \beta^\top}(\beta)\, \text{V}\, \frac{\delta h}{\delta \beta}(\beta)^\top\right)^{-1/2} \mathcal{N}\left(0, \frac{\delta h}{\delta \beta^\top}(\beta)\, \text{V}\, \frac{\delta h}{\delta \beta}(\beta)^\top\right)$$
$$= \mathcal{N}\left(0, \mathbf{I}_q\right)$$

Et par la propriété des applications continues, sous H_0,

$$W_n \xrightarrow{d} \chi^2_q$$

ce qui complète la démonstration du point 1 de la propriété.

(2) Sous l'hypothèse alternative, $h(\beta) \neq 0$, par le théorème de Slustsky,

$$h(\hat{\beta}_n) \xrightarrow{p} h(\beta) \neq 0$$

Par conséquent,

$$W_n/n \xrightarrow{p} h(\beta)^\top \left(\frac{\delta h}{\delta \beta^\top}(\beta)\, \mathrm{V}\, \frac{\delta h}{\delta \beta^\top}(\beta)^\top \right)^{-1} h(\beta)$$

et par conséquent, sous H_1,

$$W_n \to \infty$$

□

Remarquons que dans le cas de contraintes linéaires $h(\beta) = R\beta - r$, nous avons :

$$W_n = n\left(R\hat{\beta}_n - r\right)^\top \left(R\hat{\mathrm{V}}_n R^\top\right)\left(R\hat{\beta}_n - r\right)$$

En outre dans le cas homoscédastique, on peut remplacer $\hat{\mathrm{V}}_n$ par $s^2(\mathbf{X}^\top\mathbf{X}/n)^{-1}$. Alors, la statistique de Wald devient,

$$W_n = \left(R\hat{\beta}_n - r\right)^\top \left(s^2 R(\mathbf{X}^\top\mathbf{X})^{-1}R^\top\right)^{-1}\left(R\hat{\beta}_n - r\right)$$

qui est similaire à l'expression de la statistique de Fisher, mis à part l'ajustement relatif au nombre de degrés de liberté dans le numérateur.

ANNEXE A. RÉSULTATS ASYMPTOTIQUES UTILISÉS

Nous résumons les principaux résultats utilisés dans cette partie du cours en supposant connus les principaux concepts de convergence de suites de variables aléatoires. Pour une suite de variables aléatoires $\{X_n : n = 1, 2 \ldots\}$ et une variable aléatoire ou non X on note : $X_n \xrightarrow{p} X$ la convergence en probabilité, $X_n \xrightarrow{p.s.} X$ la convergence presque sûre, la $X_n \xrightarrow{d} X$ convergence en loi, $X_n \xrightarrow{L^p} X$ la convergence en moyenne d'ordre p.

Théorème A.1. ***(Loi faible des grands nombres(LFGN))*** *Soit $X_1, \ldots, X_n$ une suite de variables aléatoires i.i.d., telles que* $\mathrm{E}(|X_i|) < \infty$*. Alors, $n^{-1} \sum_{i=1}^{n} X_i \xrightarrow{p} \mathrm{E}(X_i)$ quand $n \to \infty$.*

Théorème A.2. ***(de Slutsky)*** *Supposons que $X_n \xrightarrow{p} c$, pour une constante c, et soit $h(\cdot)$ une fonction continue en c. Alors, $h(X_n) \xrightarrow{p} h(c)$.*

Théorème A.3. ***(de convergence de Cramer)***

(i) Supposons que $X_n \xrightarrow{d} X$, et $Y_n \xrightarrow{p} c$. alors,

  (a) $X_n + Y_n \xrightarrow{d} X + c$.

  (b) $Y_n X_n \xrightarrow{d} cX$,

  (c) $X_n / Y_n \xrightarrow{d} X/c$, pour $c \neq 0$.

  Remarque *: des résultats similaires sont établis pour des vecteurs/matrices avec les définitions appropriés pour la multiplication et division de vecteurs/matrices.*

(ii) Si $X_n \xrightarrow{p} X$, alors $X_n \xrightarrow{d} X$. L'inverse n'est pas vrai :

(iii) Si $X_n \xrightarrow{d} C$, une constante, alors $X_n \xrightarrow{p} C$.

(iv) Si $X_n - Y_n \xrightarrow{p} 0$, et $Y_n \xrightarrow{d} Y$, alors $X_n \xrightarrow{d} Y$.

ANNEXE B. CONVERGENCE DE L'ESTIMATEUR DE LA MATRICE DES VARIANCES-COVARIANCES(SUITE)

Dans cette annexe nous montrons que les termes $R_{1,n}$ et $R_{2,n}$ de l'équation (4.1) convergent en probabilité vers zéro. La démonstration utilise le résultat suivant appelé *inégalité de Holder*.

Propriété B.1. ***(Inégalité de Hölder)*** *Soit X et Y deux variables aléatoires. Si $p > 1$, $q > 1$, $1/p + 1/q = 1$, alors* $\mathrm{E}(|XY|) \leq (\mathrm{E}\,|X|^p)^{1/p}\,(\mathrm{E}\,|Y|^q)^{1/q}$.

Remarque : Pour $p = q = 2$ nous avons l'inégalité de Cauchy-Schwartz.

La convergence en probabilité vers zéro élément par élément est équivalente à la convergence en probabilité des normes vers zéro. La norme d'une matrice A est donnée par,

$$\begin{aligned}\|A\| &= \left(\mathrm{Tr}(A^\top A)\right)^{1/2}\\ &= \left(\sum_i \sum_j a_{ij}^2\right)^{1/2}\end{aligned}$$

où a_{ij} est l'élément (i, j) de la matrice A. Pour $R_{1,n}$,

$$\begin{aligned}\left\| n^{-1} \sum_{i=1}^n \left((\hat\beta_n - \beta)^\top X_i U_i\right) X_i X_i^\top \right\| &\leq n^{-1} \sum_{i=1}^n \left\| \left((\hat\beta_n - \beta)^\top X_i U_i\right) X_i X_i^\top \right\|\\ &= n^{-1} \sum_{i=1}^n \mathrm{Tr}\left(U_i^2 \left(\left(\hat\beta_n - \beta\right)^\top X_i\right)^2 X_i X_i^\top X_i X_i^\top\right)^{1/2}\\ &= n^{-1} \sum_{i=1}^n |U_i| \left|\left(\hat\beta_n - \beta\right)^\top X_i\right| \|X_i\|\, \mathrm{Tr}(X_i X_i^\top)^{1/2}\\ &= n^{-1} \sum_{i=1}^n |U_i| \left|\left(\hat\beta_n - \beta\right)^\top\right| \|X_i\|^2\end{aligned}$$

$$\left|(\hat\beta_n - \beta)^\top X_i\right| \leq \left\|\hat\beta_n - \beta\right\| \|X_i\|$$

Par conséquent,

$$\left\|R_{1,n}\right\| \leq \left\|\hat\beta_n - \beta\right\| n^{-1} \sum_{i=1}^n |U_i|\, \|X_i\|^3$$

Par l'inégalité de Holder avec $p = 4$ et $q = 4/3$,

$$\begin{aligned}\mathrm{E}\left(|U_i|\, \|X_i\|^3\right) &\leq (\mathrm{E}(|U_i|^4))^{1/4} (\mathrm{E}(\|X_i\|^4))^{3/4}\\ &< \infty\end{aligned}$$

étant donné que par l'hypothèse C6 nous avons $\mathrm{E}(|U_i|)^4 < \infty$, et,

$$\begin{aligned} \mathrm{E}(\|X_i\|^4) &= \mathrm{E}\left(\sum_{j=1}^{K} X_{i,j}^2\right)^2 \\ &= \sum_{j=1}^{K}\sum_{k=1}^{K} \mathrm{E}(X_{i,j}^2 X_{i,k}^2) \end{aligned} \tag{B.1}$$

où $Er(X_{i,j}^2 X_{i,k}^2) < \infty$ en raison de l'hypothèse C5, comme cela a été montré dans la propriété 3.1. Par conséquent, par la LFGN,

$$n^{-1}\sum_{i=1}^{n} |U_i|\,\|X_i\|^3 \xrightarrow{p} \mathrm{E}(|U_i|\,\|X_i\|^3)$$

et comme nous avons $\left\|\hat{\beta}_n - \beta\right\| \xrightarrow{p} 0$, nous avons que $R_{1,n} \xrightarrow{p} 0$.
Considérons maintenant le cas de $R_{2,n}$. Par des arguments similaires aux précédents, nous pouvons borner $R_{2,n}$ par,

$$\begin{aligned} \left\| n^{-1}\sum_{i=1}^{n} \left((\hat{\beta}_n - \beta)^\top X_i\right)^2 X_i X_i^\top \right\| &\leq n^{-1}\sum_{i=1}^{n} \left((\hat{\beta}_n - \beta)^\top X_i\right)^2 \|X_i\|\,\mathrm{Tr}(X_i X_i^\top)^{1/2} \\ &= \left\|(\hat{\beta}_n - \beta)\right\|^2 n^{-1}\sum_{i=1}^{n} \|X_i\|^4 \end{aligned}$$

Et par (B.1) et la LFGN,

$$n^{-1}\sum_{i=1}^{n} \|X_i\|^4 \xrightarrow{p} \mathrm{E}(\|X_i\|^4)$$

et par conséquent, $R_{2,n} \xrightarrow{p} 0$.