Master MIASH C2ES, 1ère année: Économétrie 2

Propriétés asymptotiques de l' estimateur des moindres carrés

Michal Urdanivia, Université de Grenoble Alpes, Faculté d'Économie, GAEL
Courriel: michal.wong-urdanivia@univ-grenoble-alpes.fr
Année universitaire 2016-2017

1. Introduction

Dans ces notes nous allons nous intéresser aux propriétés asymptotiques de l'estimateur des moindres carrés ordinaires(MCO par la suite) pour le modèle de régression linéaire, lequel a été longuement étudié dans le cours d'économétrie du semestre 1. Rappelons qu'il s'agit de l'estimateur du vecteur $K \times 1$, β, dans le modèle où la relation entre la variable dépendante Y et le vecteur $K \times 1$ de régresseurs X est linéaire par rapport à β,

$$Y = X^\top \beta + U$$

où U est le terme d'erreur du modèle. L'estimateur des MCO peut alors s'écrire,

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

pour un échantillon $\{(Y_i, X_i), i = 1, ..., n\}$ de Y et de X. Observez que nous indiçons l'estimateur par rapport à la taille de l'échantillon car nous allons étudier les propriétés de $\widehat{\beta}_n$ pour n "devenant de plus en plus grand". Autrement dit les propriétés asymptotiques de $\widehat{\beta}_n$.

2. 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.

Hypothèse H1. *Les données $\{(Y_i, X_i), i = 1, ..., n\}$ sont un échantillon i.i.d.*

Hypothèse H2. *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 $\mathbb{E}(U_i) = 0$.

Hypothèse H3. *X_i est (faiblement)exogène par rapport à U_i,*

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

Hypothèse H4. *La matrice $\mathbb{E}(X_i X_i^\top)$ est finie et définie positive.*

Hypothèse H5. $\mathbb{E}(X_{i,k}^4) < \infty$ *, pour tout* $k = 1, ..., K$*.*

Hypothèse H6. $\mathbb{E}(U_i^4) < \infty$

Hypothèse H7. $\mathbb{E}(U_i^2 X_i X_i^\top)$ *est définie positive.*

3. Convergence

L'estimateur des MCO est convergent pour β si $\widehat{\beta}_n \xrightarrow{p} \beta$, ce qui est établi par le théorème suivant.

Théorème 1. ***(Convergence de l'estimateur des moindres carrés)*** *Sous les hypothèses H1 - H4,* $\widehat{\beta}_n \xrightarrow{p} \beta$*.*

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

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

Les termes U_i's et les termes $X_i U_i$'s sont i.i.d. sous l'hypothèse H1. Dans ce cas, par la loi faible de grands nombres,

$$n^{-1} \sum_{i=1}^n X_i U_i \xrightarrow{p} \mathbb{E}(X_i U_i) = 0$$

Où l'on utilise H3. Dans la mesure où $E(X_i X_i^\top)$ est finie sous l'hypothèse H4 nous avons par la loi faible des grand nombres,

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

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

$$\left(n^{-1} \sum_{i=1}^n X_i X_i^\top \right)^{-1} \xrightarrow{p} \left(\mathbb{E}(X_i X_i^\top) \right)^{-1} \tag{2}$$

Et par conséquent,

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

et donc,

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

□

4. Distribution asymptotique

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

Théorème 2. ***(Normalité asymptotique)*** *Sous les hypothèses H1-H7,*

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

où

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

Démonstration. Nous avons en utilisant (1),

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

En raisons de l'hypothèse H3, $\mathbb{E}(X_i U_i) = 0$. En outre par l'inégalité de Cauchy-Schwartz et sous les hypothèses H5 , H6, et H7, nous avons pour l'élément (j,k), $j,k = 1,...,K$, de $\mathbb{V}(X_i U_i) = \mathbb{E}(U_i^2 X_i X_i^\top)$, soit $\mathbb{E}(U_i^2 X_{i,j} X_{i,k})$,

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

Par le théorème central-limite,

$$n^{-1/2}\sum_{i=1}^{n} X_i U_i \xrightarrow{d} \mathcal{N}\left(0, \mathbb{E}(U_i^2 X_i X_i^\top)\right) = \mathcal{N}(0, \Omega) \tag{3}$$

Finalement (2), (3) 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 du théorème 2 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 *: $\mathbb{E}(U_i^2|X_i) = \sigma^2(X_i)$ pour une fonction $\sigma^2 : \mathbb{R}^K \mapsto \mathbb{R}$.*

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

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

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

$$V = \sigma^2\left(\mathbb{E}(X_i X_i^\top)\right)^{-1} \tag{4}$$

En effet par la règle des conditionnements successifs,

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

ainsi dans ce cas,

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

5. Estimation de la matrice des variances-covariances

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

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

où,

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

Nous avons déjà montré que $\widehat{Q}_n^{-1} \xrightarrow{d} Q^{-1}$(c.f., (2)). Considérons maintenant $\widehat{\Omega}_n$. Nous pouvons écrire ici,

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

Il en résulte que,

$$n^{-1}\sum_{i=1}^{n}\widehat{U}_i^2 X_i X_i^\top = n^{-1}\sum_{i=1}^{n} U_i^2 X_i X_i^\top - 2R_{1,n} + R_{2,n} \tag{5}$$

où,

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

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

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

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,

$$\widehat{V}_n \xrightarrow{p} V$$

L'estimateur de la matrice des variances-covariances $\widehat{V}_n = \widehat{Q}_n^{-1}\widehat{\Omega}_n\widehat{Q}_n^{-1}$, 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 [?])

6. 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, $k = 1, ..., K$,

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

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

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

et par conséquent,

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

ce qui peut aussi s'écrire comme,

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

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

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

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

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

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.

7. 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*,

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

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

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

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

Théorème 3. *Sous les hypothèses H1-H6,*

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

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

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

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

En outre,

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

Par le théorème de convergence de Cramer, sous H_0,

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

Et par le théorème des applications continues, sous H_0,

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

ce qui complète la démonstration du point 1 du théorème.

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

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

Par conséquent,

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

et par conséquent, sous H_1,

$$W_n \to \infty$$

$\square$

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

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

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

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

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. 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 (5) convergent en probabilité vers zéro. La démonstration utilise le résultat suivant appelé *inégalité de Holder*.

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

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 $R_{1,n}$,

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

Par l'inégalité de Cauchy-Schwartz,

$$\left|(\widehat{\beta}_n-\beta)^\top X_i\right| \leq \left\|\widehat{\beta}_n-\beta\right\|\|X_i\|$$

Par conséquent,

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

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

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

étant donné que par l'hypothèse H6 nous avons $\mathbb{E}(|U_i|)^4 < \infty$, et,

$$\begin{aligned}\mathbb{E}(\|X_i\|^4) &= \mathbb{E}\left(\sum_{r=1}^K X_{i,r}^2\right)^2\\ &= \sum_{r=1}^K\sum_{s=1}^K \mathbb{E}(X_{i,r}^2X_{i,s}^2) \qquad (6)\end{aligned}$$

où $Er(X_{i,r}^2 X_{i,s}^2) < \infty$ en raison de l'hypothèse H5, comme cela a été montré dans le théorème 2. Par conséquent, par la LFGN,

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

et comme nous avons $\left\|\widehat{\beta}_n - \beta\right\| \xrightarrow{p} 0$, 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,

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

Et par (6) et la LFGN,

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

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

Références

Halbert White. *Asymptotic Theory for Econometricians.* The MIT Press, 1980.

