

Estimation dans un modèle d'équations estimantes avec données fonctionnelles

Title: An estimation procedure in a model defined by estimating equations with functional data

Matthieu Saumard¹

Résumé : Nous proposons un nouveau modèle impliquant des données fonctionnelles. Il s'agit d'un modèle d'équations estimantes conditionnelles que l'on peut aussi appeler modèle de restrictions de moments conditionnels. Nous proposons un estimateur consistant du paramètre d'intérêt. La convergence en loi de cet estimateur est aussi démontrée. Nous étudions ensuite la convergence uniforme des instruments optimaux de ce modèle.

Abstract: We consider a conditional moment restrictions model for functional data. The response variable is a vector in a finite dimensional space while the conditioning variable as well as the parameter to be estimated are elements of a functional space. To estimate the parameter we first decompose the conditioning variable and the parameter in a basis of the functional space. For a given sample size we truncate their decompositions to a finite number of coefficients and we apply the classical approach for estimating equations. For the asymptotic results we let this finite number of coefficients to grow to infinity at a suitable rate.

Mots-clés : Données fonctionnelles, Equations estimantes, G.M.M.

Keywords: Functional data, Estimating equations, G.M.M.

Classification AMS 2000 : 62J02, 62J12, 62G08

1. Introduction

Plusieurs modèles économétriques impliquent des restrictions de moments conditionnels. Hansen (1982) a proposé pour estimer le paramètre d'intérêt dans ce type de modèle la méthode des moments généralisée (méthode G.M.M.). Ces modèles ont été étudiés par la suite par Newey (1993). Plusieurs autres méthodes, comme par exemple la vraisemblance empirique, ont été proposées depuis ; citons Kitamura et al. (2004), Smith (2007) et Domínguez and Lobato (2004). Toutes ces méthodes s'appliquent à des variables finies-dimensionnelles.

Le modèle étudié le plus proche de celui que l'on considère est celui de Müller and Stadtmüller (2005). Ils étudient les modèles linéaires fonctionnels généralisés. Chen et al. (2011) propose un 'single index' modèle pour des données fonctionnelles avec une fonction de lien non-paramétrique. Dans un sens, cette approche englobe celle de Müller and Stadtmüller (2005). Notre modèle est aussi une généralisation du modèle de Müller and Stadtmüller (2005). L'idée principale est de réduire la dimension des variables fonctionnelles en les décomposant dans une base orthonormale et d'utiliser les p premières composantes pour se ramener à un modèle fini-dimensionnel, puis de faire tendre $p = p_n$ vers l'infini. Nous allons utiliser la même idée de réduction de la dimension. Cependant, la méthode d'estimation diffère de celle de Müller and Stadtmüller (2005) et

¹ IRMAR-INSA & Instituto de Estadística de Pontificia Universidad Católica de Valparaíso (Chile)
E-mail : matthieu.saumard@gmail.com

l'estimateur obtenu est différent quand les modèles coïncident. En effet, on utilise la méthode des moments généralisée pour estimer les p premières valeurs du paramètre d'intérêt fonctionnel dans la décomposition dans la base orthonormale. Cette fonction paramètre d'intérêt remplace le vecteur paramètre d'intérêt dans les modèles finis-dimensionnels de restrictions de moments conditionnels.

Nous obtenons un résultat asymptotique de consistance de l'estimateur vers la fonction paramètre d'intérêt, quand la dimension de troncature croît avec la dimension de l'échantillon. Le deuxième résultat est un résultat asymptotique limite (type théorème centrale limite) de la déviation entre le paramètre fonctionnel estimé et le vrai paramètre fonctionnel.

L'article est organisé comme suit : la description du modèle, des notations et des considérations préliminaires se trouvent dans la Section 2. La procédure d'estimation du paramètre fonctionnel est décrite dans la Section 3. Les résultats asymptotiques sont regroupés dans la Section 4. Dans la Section 5, nous estimons les instruments optimaux dans un cas particulier. Une étude de simulation est présentée dans la Section 6. L'article se termine par une brève conclusion, Section 7. Enfin, les preuves des résultats sont détaillées à la fin de l'article, Section 8.

2. Le modèle

On dispose de n observations i.i.d. $z_i = (x_i, y_i)$, $i = 1, \dots, n$ issues du couple (x, y) où $x \in L^2[0, 1]$ est une courbe aléatoire et y est une variable aléatoire réelle. Au lieu d'utiliser l'espace $L^2[0, 1]$, on peut utiliser un espace plus général, comme un espace de Hilbert séparable. On notera g une fonction connue à valeurs vectorielles de dimension $s \times 1$. Notre but est d'estimer un paramètre d'intérêt $\theta_0 \in L^2[0, 1]$ défini par

$$\mathbb{E}[g(z, \theta) | x] = 0 \iff \theta = \theta_0. \quad (2.1)$$

Nous allons nous cantonner au cas :

$$g((x, y), \theta) = \bar{g} \left(y, \int_0^1 x(t) \theta(t) dt \right).$$

Le modèle linéaire généralisé fonctionnel étudié par Müller and Stadtmüller (2005) peut s'écrire de la manière suivante :

$$y_i = \rho \left(\alpha + \int_0^1 \theta(t) x_i(t) dt \right) + e_i, \quad i = 1 \dots n, \quad (2.2)$$

où ρ est une fonction à valeurs réelles et le terme d'erreur satisfait les conditions $\mathbb{E}[e_i | x] = 0$ et $\text{Var}(e_i | x) = \sigma^2(\mu)$, et $\mu = \rho \left(\alpha + \int_0^1 \theta(t) x(t) dt \right)$. Le modèle (2.1) est une généralisation du modèle (2.2). En posant $\bar{g}(y, \int_0^1 x(t) \theta(t) dt) = y - \rho \left(\alpha + \int_0^1 \theta(t) x(t) dt \right)$, on retrouve les restrictions de moments conditionnels : $\mathbb{E} \left[\bar{g}(y, \int_0^1 x(t) \theta(t) dt) | x \right] = 0$. Le modèle (2.1) que nous étudions ici a donc aussi comme cas particulier le modèle linéaire fonctionnel. En effet, il est déjà un cas particulier du modèle linéaire généralisé fonctionnel (2.2) ; il suffit de prendre $\rho(z) = z$, $\forall z \in \mathbb{R}$ et ainsi nous obtenons

$$y_i = \alpha + \int_0^1 \theta(t) x_i(t) dt + e_i, \quad i = 1 \dots n,$$

qui est le modèle linéaire fonctionnel.

Nous pouvons d'ores et déjà faire une remarque pour comparer les deux approches. Le modèle d'étude (2.1) va être étudié par une approche G.M.M., elle diffère de la procédure d'estimation utilisée par Müller et Stadtmüller (2005). En particulier, l'estimateur obtenu ne coïncide pas avec l'estimateur de Müller et Stadtmüller (2005). Ainsi, notre procédure est aussi une nouvelle procédure d'estimation pour les modèles linéaires fonctionnels généralisés.

Un troisième exemple de modèle inclus dans le modèle d'équations estimantes (2.1) est le modèle de régression quantile. De plus, ce dernier modèle ne s'écrit pas sous la forme du modèle linéaire généralisé fonctionnel. Le modèle de régression quantile s'écrit :

$$\mathbb{P}\left(y \leq \int_0^1 x(t)\theta(t)dt \mid x\right) = \tau.$$

En posant $g(y, z) = 1_{\{y \leq z\}} - \tau$, nous retrouvons les restrictions de moments conditionnels

$$\mathbb{E}\left[g(y, \int_0^1 x(t)\theta(t)dt) \mid x\right] = 0.$$

Il est à noter que notre méthodologie d'estimation, ainsi que les hypothèses du théorème 4.1, ne nous permettent pas d'inclure ce modèle. Nous ajoutons donc un exemple important qui peut s'inclure dans notre méthodologie.

Un autre exemple intéressant est le modèle fonctionnel à variable endogène factice, voir Newey (1993) pour une présentation plus détaillée. C'est un modèle important en économie qui traite de l'effet d'un événement s sur une variable économique, comme par exemple l'effet de l'éducation sur le salaire. Le modèle peut s'écrire :

$$y = \alpha_0 s + f\left(\int_0^1 x(t)\beta_0(t)dt\right) + \varepsilon, \quad s \in \{0, 1\}, \quad \mathbb{E}[\varepsilon \mid x] = 0.$$

Par une modification mineure, notre procédure permet d'inclure ce modèle, qui clairement n'est pas inclus dans le modèle linéaire fonctionnel généralisé.

Soit ρ_j , $j = 1, 2, \dots$, une base orthonormale de l'espace $L^2[0, 1]$, on a donc $\int_0^1 \rho_j(t)\rho_k(t)dt = \delta_{jk}$. Ainsi $x(t)$ et $\theta(t)$ peuvent se décomposer en :

$$x(t) = \sum_{j=1}^{\infty} \varepsilon_j \rho_j(t), \quad \theta(t) = \sum_{j=1}^{\infty} \theta_j \rho_j(t),$$

l'égalité étant prise au sens des fonctions $L^2[0, 1]$, c'est-à-dire presque sûrement. Les coefficients ε_j et θ_j sont donnés par $\varepsilon_j = \int x(t)\rho_j(t)dt$ et par $\theta_j = \int \theta(t)\rho_j(t)dt$. De l'orthonormalité de la base ρ_j , il vient immédiatement que

$$\int_0^1 x(t)\theta(t)dt = \sum_{j=1}^{\infty} \varepsilon_j \theta_j.$$

On tronque alors cette somme aux $p = p_n$ premiers termes pour pouvoir se ramener à un modèle fini-dimensionnel. On note $x^{(p)}$ le vecteur de taille p et ε_j sa j -ème composante (notation générique pour désigner la troncation d'un élément de $L^2[0, 1]$ en un vecteur de dimension p) :

$$x^{(p)} = \left(\varepsilon_1 = \int_0^1 x(t)\rho_1(t)dt, \dots, \varepsilon_p = \int_0^1 x(t)\rho_p(t)dt \right)'.$$

Nous rencontrons alors un problème d'approximation de modèle. En effet, nous ne sommes pas sûr de trouver $\mathbb{E} \left[g(y, \sum_{j=1}^p \varepsilon_j \theta_{0,j}) \mid x^{(p)} \right] = 0$. Cependant, quand $p = p_n \rightarrow +\infty$, nous pouvons écrire, sous de bonnes conditions :

$$\mathbb{E} \left[g(y, \sum_{j=1}^p \varepsilon_j \theta_{0,j}) \mid x^{(p)} \right] = O_{\mathbb{P}}(r_n), \quad (2.3)$$

où r_n est un nombre réel qui tend vers zéro, qu'on définit ci-dessous. En effet, posons $U_p^0 = \sum_{j=1}^p \varepsilon_j \theta_{0,j}$ et $V_p^0 = \sum_{j=p+1}^{\infty} \varepsilon_j \theta_{0,j}$ de tel sorte qu'on ait $\int_0^1 x(t)\theta_0(t)dt = U_p^0 + V_p^0$, on obtient alors

$$\begin{aligned} \mathbb{E} \left[g(y, U_p^0) \mid x^{(p)} \right] &= \mathbb{E} \left[g(y, U_p^0) - \mathbb{E}[g(y, U_p^0 + V_p^0) \mid x] \mid x^{(p)} \right] \\ &= \mathbb{E} \left[g(y, U_p^0) - g(y, U_p^0 + V_p^0) \mid x^{(p)} \right] \\ &= \int g(s, U_p^0) - g(s, U_p^0 + V_p^0) dF_{y|x^{(p)}}(s). \end{aligned}$$

Sous les conditions $|g(s, u_1) - g(s, u_2)| \leq \Phi(s)|u_1 - u_2|$ et $\int \Phi(s) dF_{y|x^{(p)}}(s) \leq C, \forall p$, pour une constante C , nous pouvons majorer l'intégrande :

$$\begin{aligned} \mathbb{E} \left(\mathbb{E}^2 \left[g(y, U_p^0) \mid x^{(p)} \right] \right) &\leq \mathbb{E} \left[\left(\int \Phi(s) dF_{y|x^{(p)}}(s) \right)^2 (V_p^0)^2 \right] \\ &\leq C^2 \mathbb{E}[(V_p^0)^2] \\ &\leq C^2 \sum_{j=p+1}^{\infty} \theta_{0,j}^2 \sum_{j=p+1}^{\infty} \sigma_j^2 \end{aligned}$$

où $\sigma_j^2 = \mathbb{E}[\varepsilon_j^2]$. On peut donc prendre $r_n = \sqrt{\sum_{j=p+1}^{\infty} \theta_{0,j}^2 \sum_{j=p+1}^{\infty} \sigma_j^2}$. L'erreur d'approximation est directement reliée à $Var(V_p^0)$ et est contrôlée par la suite des variances de ε_j , et la valeur du vrai paramètre. La base ρ_j joue ici un rôle primordial, plus les premiers coefficients dans la décomposition sont informatifs, plus l'erreur d'approximation sera faible. Il est à noter que des simplifications apparaissent quand la base choisie est celle de Karhunen-Loève pour x : $\mathbb{E}[(V_p^0)^2]$ devient alors $\sum_{j \geq p+1} \theta_0^2 \sigma_j^2$, terme d'erreur d'approximation qui est optimal dans le modèle linéaire fonctionnel.

3. Estimation

Avant de pouvoir mettre en oeuvre la méthode, il faut ramener les restrictions de moments conditionnels à des restrictions de moments inconditionnels. On utilise pour cela une variable

instrumentale $A(x)$ de dimension $r \times s$. De

$$\mathbb{E} \left[g \left(y, \int x \theta_0 \right) \mid x \right] = 0,$$

on peut écrire

$$\mathbb{E} \left[A(x) g \left(y, \int x \theta_0 \right) \right] = 0.$$

Pour plus de simplicité, on utilise les instruments optimaux $B(x^{(p)})$ et dans un premier temps, on considère ces instruments optimaux connus. Pour rappel, les instruments optimaux sont ceux qui minimisent la variance asymptotique. Dans cette partie, nous avons juste besoin de connaître la définition des instruments optimaux. Les propriétés, ainsi que quelques exemples, sont proposés dans la section 5. Rappelons la définition des instruments optimaux :

$$\begin{aligned} B(x^{(p)}) &= D(x^{(p)})' \Omega^{-1}(x^{(p)}) \quad \text{où} \\ D(x^{(p)}) &= \mathbb{E} \left[\frac{\partial g}{\partial \theta} (y, x^{(p)}, \theta_0^{(p)}) \mid x^{(p)} \right] \quad \text{et} \\ \Omega(x^{(p)}) &= \mathbb{E} [g(y, x^{(p)}, \theta_0^{(p)}) g(y, x^{(p)}, \theta_0^{(p)})' \mid x^{(p)}]. \end{aligned}$$

$D(x^{(p)})$ est une matrice de taille $s \times p$, $\Omega(x^{(p)})$ est une matrice de taille $s \times s$, et la matrice des instruments optimaux $B(x^{(p)})$ est donc de taille $p \times s$. Dans notre notation de la dérivée, nous faisons un abus de notation : en fait

$$\frac{\partial g}{\partial \theta} (y, x^{(p)}, \theta_0^{(p)}) = x'^{(p)} g'_2(y, \langle x^{(p)}, \theta_0 \rangle),$$

dans le cas où $s = 1$ et g'_2 est la dérivée de g par rapport à la deuxième variable. Nous faisons aussi un abus de notation dans l'expression $g(y, x^{(p)}, \theta_0^{(p)})$ qui est égale à $g(y, \langle x^{(p)}, \theta_0^{(p)} \rangle)$. Les instruments optimaux sont évidemment inconnus, puisqu'ils dépendent de θ_0 , mais nous les utilisons dans l'analyse statistique comme des quantités connues. Nous verrons dans la section 5 comment estimer ces quantités et les remplacer dans l'estimateur. Soit \hat{P} une matrice aléatoire $p \times p$ semi-définie positive qui peut-être dans un premier temps égale à l'identité. Considérons l'estimateur

$$\hat{\theta}_n = \arg \min_{\theta \in \Theta} \hat{g}_n(\theta)' \hat{P} \hat{g}_n(\theta)$$

où

$$\hat{g}_n(\theta) = \frac{1}{n} \sum_{i=1}^n B(x_i^{(p)}) g(y_i, \sum_{j=1}^p \theta_j \varepsilon_j^i).$$

Et ε_j^i représente la j -ème coordonnée de x_i dans la base orthonormée. $\hat{\theta}_n$ est un vecteur de dimension p , que nous comparons à

$$\theta_0^{(p)} = \left(\theta_{0,1} = \int_0^1 \theta_0(t) \rho_1(t) dt, \dots, \theta_{0,p} = \int_0^1 \theta_0(t) \rho_p(t) dt \right)'$$

C'est l'objet de la section suivante.

4. Résultats asymptotiques

Introduisons tout d'abord quelques notations : on note $\|\cdot\|$ les normes quadratiques des vecteurs de \mathbb{R}^p (bien qu'elles varient avec n), on note de plus $\xi_{k_1, k'_1} = \Lambda_{k_1, k'_1}$, où Λ est la matrice $(\mathbb{E}[D(x)' \Omega(x)^{-1} D(x)])^{-1}$ de taille $p \times p$; $\xi_{k_1, k'_1}^{(1/2)} = \Lambda_{k_1, k'_1}^{-1/2}$. Pour alléger les notations, on écrit $\tilde{g}^k = B(x_k^{(p)})g(y_k, \sum_{j=1}^p \theta_j \varepsilon_j^k)$, qui est un vecteur de taille p , de plus \tilde{g}_k est la k -ème coordonnée du vecteur \tilde{g}^1 .

Nous sommes maintenant en mesure d'établir un résultat de consistance :

Lemme 4.1. *Sous les conditions*

(a) Θ est compact.

(b) soit Φ une fonction telle que pour tout $s \in \mathbb{R}$, $g(s, \cdot)$ est Lipschitzienne :

$$|g(s, u_1) - g(s, u_2)| \leq \Phi(s)|u_1 - u_2|.$$

(c) la fonction Φ définie en (b) vérifie pour une constante $C > 0$,

$$\int \Phi(s) dF_{y|x}(s) \leq C.$$

(d) condition d'uniformité sur les troncatures :

$$\sup_p \mathbb{E} \sup_{\theta} \|B(x^{(p)})g(y, \langle x^{(p)}, \theta^{(p)} \rangle)\| < \infty.$$

on a

$$\|\hat{\theta}_n - \theta_0^{(p)}\| = o_{\mathbb{P}}(1).$$

Ce résultat porte sur la consistance de l'estimateur sur les coordonnées de $\theta_0(\cdot)$ dans la base ρ_j . Il ne s'agit pas d'un résultat sur l'estimateur fonctionnel vers la vraie fonction $\theta_0(\cdot)$. Nous nous appuyons sur la démonstration classique de la consistance de l'estimateur G.M.M. pour démontrer ce résultat. Les hypothèses données ici sont cependant plus fortes que dans le cas classique. Par exemple, pour l'hypothèse (d), on a besoin d'une uniformité en p , ce qui nous permet d'appliquer une loi des grands nombres uniforme.

Donnons quelques hypothèses supplémentaires pour établir le résultat central :

Hypothèses D.

(a) La fonction connue g est deux fois dérivable.

(b) La fonction g vérifie pour un $a \in]0, 1]$

$$\left\| \frac{\partial g}{\partial \theta}(z_i, b) - \frac{\partial g}{\partial \theta}(z_i, d) \right\|_2 \leq c(y_i) \|b - d\|^a \quad \forall i = 1, \dots, n, \forall b, d,$$

et

$$\|B_j(x_i)\| |c(y_i)| \leq \Gamma(z_i), \forall j = 1, \dots, p, \text{ avec } \mathbb{E}[\Gamma(z_i)] < \infty \forall i = 1, \dots, n.$$

(c) Conditions sur les coefficients : il existe α, β, γ tel que

$$\theta_{0,j} \sim j^{-\alpha}, \quad \sigma_j \sim j^{-\beta}, \quad p = n^\gamma,$$

avec $\alpha \geq 2, \beta \geq 2, 0 < \gamma < 1/4$ et

$$1 - 2\gamma(\alpha + \beta - 1) + \frac{3}{2}\gamma < 0.$$

(d) on a

$$\sum_{k_1, k_2, k'_1, k'_2}^p \xi_{k_1, k'_1} \xi_{k_2, k'_2} \mathbb{E} \left[\left(\frac{\partial \tilde{g}^k}{\partial \theta} \right)_{k_1, k_2} \left(\frac{\partial \tilde{g}^k}{\partial \theta} \right)_{k'_1, k'_2} \right] = o(n/p^2) \quad , \forall k.$$

(e) on a

$$\sum_{k_1, k_2, k_3, k_4=1}^p \mathbb{E}[\tilde{g}_{k_1} \tilde{g}_{k_2} \tilde{g}_{k_3} \tilde{g}_{k_4}] \xi_{k_1, k_2} \xi_{k_3, k_4} = o(n/p^2).$$

(f) on a

$$\sum_{k_1, \dots, k_8=1}^p \mathbb{E}[\tilde{g}_{k_1} \tilde{g}_{k_3} \tilde{g}_{k_5} \tilde{g}_{k_7}] \mathbb{E}[\tilde{g}_{k_2} \tilde{g}_{k_4} \tilde{g}_{k_6} \tilde{g}_{k_8}] \xi_{k_1, k_2} \xi_{k_3, k_4} \xi_{k_5, k_6} \xi_{k_7, k_8} = o(n^2 p^2).$$

Commentons ces hypothèses. Les conditions sur les coefficients nous permettent d'écrire $p = o(n^{1/4})$, $r_n \sim n^{-\gamma(\alpha+\beta-1)}$ et ainsi on obtient la condition $nr_n^2 p^{3/2} = o(1)$. Cette dernière condition est utilisée dans les preuves et est valide pour $\alpha = \beta = 2$ et $\gamma = 1/4$. Les hypothèses **D**–(e) – (f) dans le cas du modèle linéaire fonctionnel généralisé décrit par Müller and Stadtmüller (2005) sont exactement les mêmes. C'est-à-dire que si on pose $\tilde{g}(y, \int_0^1 x(t)\theta(t)dt) = y - \rho \left(\alpha + \int_0^1 \theta(t)x(t)dt \right)$, alors on retrouve les conditions M–(3) et M–(4) du théorème 4.1 de Müller and Stadtmüller (2005). L'hypothèse **D**–(d) s'ajoute du fait que l'on considère un modèle d'équations estimantes. Nous pouvons maintenant formuler le résultat principal :

Théorème 4.1. *Sous les hypothèses **D** et sous les conditions du lemme 4.1, et en posant Λ tel que $\Lambda = (\mathbb{E}[D(x)' \Omega(x)^{-1} D(x)])^{-1}$, on obtient :*

$$\frac{n \left(\hat{\theta}_n - \theta_0^{(p)} \right)' \Lambda^{-1} \left(\hat{\theta}_n - \theta_0^{(p)} \right) - p}{\sqrt{2p}} \Rightarrow N(0, 1)$$

Notons

$$\hat{\theta}_n(t) = \sum_{i=1}^p \hat{\theta}_{n,i} \rho_i(t),$$

l'estimateur fonctionnel de θ_0 . Pour pouvoir comparer ces deux fonctions, il faut une distance dans $L^2[0, 1]$. La distance usuelle ne convient pas, puisque la matrice Λ^{-1} intervient dans le théorème (4.1) sous la forme $\left(\hat{\theta}_n - \theta_0^{(p)} \right)' \Lambda^{-1} \left(\hat{\theta}_n - \theta_0^{(p)} \right)$ qui n'est pas la distance usuelle de \mathbb{R}^p . Nous allons construire une distance, à l'aide d'un noyau G , qui résout ce problème. Soit, pour une fonction g réelle c'est-à-dire $s = 1$,

$$G(u, t) = \mathbb{E} \left[\frac{\mathbb{E}^2[g'_2(y, \int x \theta_0) | x]}{\mathbb{E}[g^2(y, \int x \theta_0) | x]} x(u)x(t) \right], \tag{4.4}$$

un noyau intégral, où la dérivée de g est prise sur la seconde variable. Sous la condition,

$$\mathbb{E} \left[\frac{\mathbb{E}^2[g'_2(y, f x \theta_0) | x]}{\mathbb{E}[g^2(y, f x \theta_0) | x]} \|x\|^2 \right] < +\infty,$$

G est défini presque partout sur $[0, 1]^2$. En effet, il suffit pour cela de montrer que G est dans $L^1([0, 1]^2)$. Nous avons

$$\begin{aligned} \iint_{[0,1]^2} |G(u,t)| dudt &\leq \iint_{[0,1]^2} \mathbb{E} \left[\frac{\mathbb{E}^2[g'_2(y, f x \theta_0) | x]}{\mathbb{E}[g^2(y, f x \theta_0) | x]} |x(u)||x(t)| \right] dudt \\ &\leq \mathbb{E} \left[\frac{\mathbb{E}^2[g'_2(y, f x \theta_0) | x]}{\mathbb{E}[g^2(y, f x \theta_0) | x]} \|x\|^2 \right] < +\infty \end{aligned}$$

Faisons correspondre à ce noyau intégral la distance dans $L^2[0, 1]$ définie comme suit :

$$d_G^2(f, g) = \iint (f(u) - g(u))(f(t) - g(t))G(u,t) dt du, \forall f, g \in L^2[0, 1].$$

Nous pouvons aussi définir l'opérateur linéaire intégral A_G correspondant à G :

$$(A_G f)(t) = \int f(u)G(u,t) du.$$

Cet opérateur est diagonalisable dans $L^2[0, 1]$ si $\int |G(u,t)|^2 dudt < \infty$. Notons $\{\rho_j^G, \lambda_j^G, j = 1, 2, \dots\}$, les fonctions propres et les valeurs propres de l'opérateur A_G . La distance d_G peut s'exprimer comme

$$d_G^2(f, g) = \sum_k \lambda_k^G (f_{\rho^{G,k}} - g_{\rho^{G,k}})^2,$$

où $f_{\rho^{G,k}}$ (respectivement $g_{\rho^{G,k}}$) est le k -ème coefficient dans la décomposition de f (respectivement g) dans la base ρ^G . Nous obtenons alors

$$d_G^2(\hat{\theta}_n(\cdot), \theta_0(\cdot)) = \left(\hat{\theta}_n^G - \theta_0^{(p)G} \right)' \Lambda^{-1} \left(\hat{\theta}_n^G - \theta_0^{(p)G} \right) + \sum_{j=p+1}^{\infty} \lambda_j^G \theta_{0,j}^2,$$

où tous les calculs sont effectués dans la base ρ^G . En effet,

$$\begin{aligned} \delta_{j,k} \lambda_j^G &= \iint G(u,t) \rho_j^G(s) \rho_k^G(t) dudt \\ &= \iint \mathbb{E} \left[x(u)x(t) \frac{\mathbb{E}^2[g'_2(y, f x \theta_0) | x]}{\mathbb{E}[g^2(y, f x \theta_0) | x]} \right] \rho_j^G(u) \rho_k^G(t) dudt \\ &= \mathbb{E} \left[\varepsilon_j \varepsilon_k \frac{\mathbb{E}^2[g'_2(y, f x \theta_0) | x]}{\mathbb{E}[g^2(y, f x \theta_0) | x]} \right] \\ &= \mathbb{E} \left[\varepsilon_j^G \varepsilon_k^G \frac{\mathbb{E}^2[g'_2(y, f x^{(p)} \theta_0) | x^{(p)}]}{\mathbb{E}[g^2(y, f x^{(p)} \theta_0) | x^{(p)}]} \right] \\ &= \Lambda_{j,k}^{-1}, \end{aligned}$$

en remplaçant les ε_j par ε_j^G qui sont donnés par

$$\varepsilon_j^G = \frac{\mathbb{E}[g'_2(y, \int x \theta_0) | x]}{\mathbb{E}[g'_2(y, \int x^{(p)} \theta_0) | x^{(p)}]} \frac{\sqrt{\mathbb{E}[g^2(y, \int x^{(p)} \theta_0) | x^{(p)}]}}{\sqrt{\mathbb{E}[g^2(y, \int x \theta_0) | x]}} \int x(t) \rho_j^G(t) dt,$$

Toutes ces considérations donnent le résultat suivant :

Corollaire 4.2. *Si la fonction $\theta_0(\cdot)$ vérifie la propriété*

$$\sum_{j=p+1} \lambda_j^G \left(\int_0^1 \theta_0(t) \rho_j^G(t) dt \right)^2 = o\left(\frac{\sqrt{p}}{n}\right), \tag{4.5}$$

alors

$$\frac{nd_G^2(\hat{\theta}_n(\cdot), \theta_0(\cdot)) - p}{\sqrt{2p}} \Rightarrow N(0, 1).$$

Commentons un peu ce Corollaire : la condition (4.5) peut se vérifier si on impose le même type de condition que l'hypothèse D(c) sur les valeurs propres et vecteurs propres de A_G . Supposons que $\theta_{0,j}^G \sim j^{-\alpha}$ et $\lambda_j^G \sim j^{-l}$, alors nous obtenons une condition du type $\gamma(2\alpha + l - 1/2) > 1$, qui peut être vérifiée par exemple quand $\alpha = 2$, $l = 2$ et $\gamma = 1/4$.

5. Les instruments optimaux

Nous nous proposons dans cette partie d'estimer

$$B(x) = D'(x)\Omega^{-1}(x)$$

où

$$D(x) = \mathbb{E} \left[\frac{\partial g}{\partial \theta} (Y, \langle X, \theta_0 \rangle) | X = x \right]$$

est une matrice de taille $s \times p$ et où

$$\Omega(x) = \mathbb{E} [g(Y, \langle X, \theta_0 \rangle)g'(Y, \langle X, \theta_0 \rangle) | X = x]$$

est une matrice de taille $s \times s$. On suppose dans la suite que $s = 1$ pour alléger les notations.

Faisons une remarque sur les instruments optimaux dans le cas du modèle linéaire fonctionnel. Les instruments optimaux sont alors connus, en effet si on se place dans le cadre du modèle linéaire fonctionnel, $Y = \int_0^1 X(t)\theta_0(t)dt + e$, alors

$$g\left(Y, \int X \theta\right) = Y - \int X \theta,$$

ainsi

$$D(x) = x$$

et

$$\Omega(x) = \mathbb{E}(e^2 | x) = \sigma^2(x),$$

où e est le terme d'erreur dans le modèle linéaire fonctionnel. Le fait de connaître les instruments optimaux dans le cas de modèle particulier renforce la partie d'estimation dans le cas d'instruments optimaux supposés connus. Il convient néanmoins d'estimer ces instruments dans le cas où ils ne sont pas connus.

Soit Φ une fonction, on peut s'intéresser au cas, plus général, de l'estimation de

$$m(x, t) = \mathbb{E}[\Phi(Y, t) | X = x].$$

K est un noyau et h est un paramètre fenêtre associé à K . Nous pouvons maintenant définir l'estimateur :

$$\hat{m}(x, t) = \frac{\sum_{i=1}^n \Phi(Y_i, t) K(h^{-1}d(X_i, x))}{\sum_{i=1}^n K(h^{-1}d(X_i, x))}, \quad (5.6)$$

où x appartient à un sous-espace compact S de $L^2[0, 1]$ pour la norme induite par $d(u, v) := \|u - v\|$ et t est un réel. Nous établissons un résultat de convergence uniforme pour \hat{m} vers m :

$$\sup_{t \in \mathbb{R}, x \in S} |\hat{m}(x, t) - m(x, t)| = o_{\mathbb{P}}(1).$$

Ainsi, en disposant d'un estimateur initial $\hat{\theta}_n$ de θ_0 , nous pouvons écrire

$$\begin{aligned} & |\hat{m}(x, \langle x, \hat{\theta}_n \rangle) - m(x, \langle x, \theta_0 \rangle)| \\ &= |\hat{m}(x, \langle x, \hat{\theta}_n \rangle) - \hat{m}(x, \langle x, \theta_0 \rangle) + \hat{m}(x, \langle x, \theta_0 \rangle) - m(x, \langle x, \theta_0 \rangle)| \\ &\leq \left\{ \sup_{t \in \mathbb{R}, x \in S} |\hat{m}(x, t) - m(x, t)| \right\} + C \|x\| \|\hat{\theta}_n - \theta_0\| \\ &\leq \left\{ \sup_{t \in \mathbb{R}, x \in S} |\hat{m}(x, t) - m(x, t)| \right\} + C' \|\hat{\theta}_n - \theta_0\|, \end{aligned}$$

où C et C' sont des constantes. Par exemple, si

$$\Phi(y, t) = \frac{\partial g}{\partial \theta}(y, t),$$

alors $D(x) = m(x, \langle x, \theta_0 \rangle)$ et son estimateur est $\hat{D}(x) = \hat{m}(x, \langle x, \hat{\theta}_n \rangle)$.

La distance $d(\cdot, \cdot)$ joue un rôle primordial. On choisit au préalable une base de $L^2[0, 1]$ que l'on note (e_j) pour $j = 1, 2, \dots$. La distance que l'on utilise est induite par la norme usuelle de $L^2[0, 1]$. Rappelons que

$$\chi^{(k)} = \left(\chi_1 = \int_0^1 \chi(t) e_1(t) dt, \dots, \chi_k = \int_0^1 \chi(t) e_k(t) dt \right)'$$

Soit S un compact de $L^2[0, 1]$ pour la distance d , définissons l'application ζ de $L^2[0, 1]$ dans \mathbb{R}^k qui à une fonction χ fait correspondre $\chi^{(k)}$, $S^{(k)} = \zeta(S)$ est alors un compact de \mathbb{R}^k pour la norme 2. Pour tout ensemble I de \mathbb{R}^k , I^ε est l'ensemble des ε -voisins pour la norme max $|\cdot|_+$, qui est $|x|_+ = \max_{1 \leq i \leq k} |x_i|$, pour $x \in \mathbb{R}^k$.

Faisons une remarque essentielle, on peut décomposer $\hat{m}(x, t) - m(x, t)$ de la façon suivante :

$$\hat{m}(x, t) - m(x, t) = \hat{m}(x, t) - \hat{m}(x^{(k)}, t) + \hat{m}(x^{(k)}, t) - m(x^{(k)}, t) + m(x^{(k)}, t) - m(x, t).$$

Ainsi, le terme $\hat{m}(x^{(k)}, t) - m(x^{(k)}, t)$ est ramené à un problème fini-dimensionnel que l'on traite à l'aide d'Einmahl and Mason (2005). Faisons maintenant quelques hypothèses.

Hypothèses E.

(a) $\forall x \in S, 0 < C\psi(h) \leq \mathbb{P}(X \in B(x, h)) \leq C'\psi(h) < \infty$, pour C, C' des constantes. ψ vérifie

$$\exists C^* > 0, \exists \varepsilon_0, \forall \varepsilon < \varepsilon_0, \int_0^\varepsilon \psi(u)du > C^*\varepsilon\psi(\varepsilon),$$

$\psi(h) \rightarrow 0$ quand $h \rightarrow 0$ et

$$\frac{1}{h\psi(h)} \sup_{x \in S} \|x - x^{(k)}\| \rightarrow 0.$$

(b) $\forall x_1, x_2 \in S, \forall t \in \mathbb{R}$,

$$|m(x_1, t) - m(x_2, t)| \leq C(t)d(x_1, x_2)$$

avec $\sup_{t \in \mathbb{R}} C(t) < +\infty$.

(c) $\forall k, X^{(k)}$ admet une densité continue et strictement positive sur $S^{(k)\varepsilon}$, pour un $0 < \varepsilon < 1$.

(d) K est positif, de support $[0, 1]$, d'intégrale 1, sa dérivée existe et vérifie la condition $-\infty < C \leq K'(t) \leq C' < 0$, pour des constantes C, C' . De plus, $nh^k \rightarrow +\infty$.

(e) L'enveloppe F est défini par $\sup_{t \in \mathbb{R}} |\Phi(y, t)| \leq F(y), y \in \mathbb{R}$. Elle vérifie

$$\forall m \geq 2, \mathbb{E}[F^m(Y)|X = x] < \delta_m(x) < C < +\infty$$

avec δ_m continue sur S et C une constante, de plus

$$\sup_k \sup_{z \in S^{(k)\varepsilon}} \mathbb{E}(F^l(Y) | X^{(k)} = z) < \infty,$$

pour un $l > 2$.

Théorème 5.1. Sous les hypothèses E, on obtient :

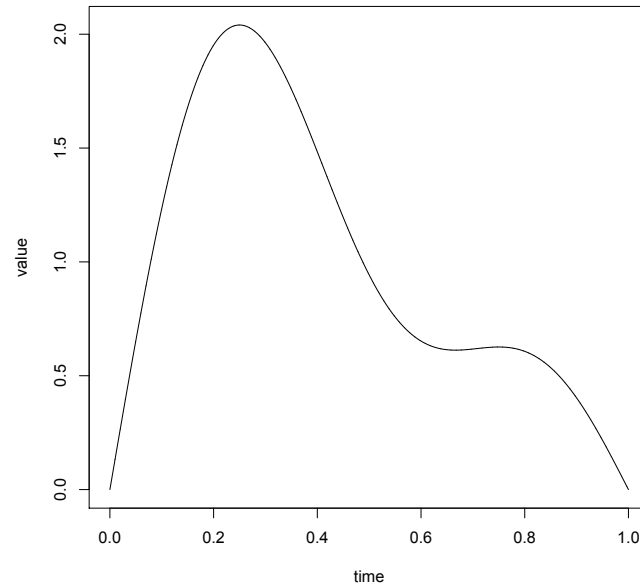
$$\sup_{t \in \mathbb{R}, x \in S} |\hat{m}(x, t) - m(x, t)| = o_{\mathbb{P}}(1).$$

Faisons quelques commentaires : comparons maintenant à l'article de Ferraty et al. (2010). La différence entre notre résultat et leur résultat se situe dans l'ajout de la variable t qui prend en compte le paramètre θ . On ne peut pas appliquer directement leur résultat à notre problématique. La condition $nh^k \rightarrow +\infty$ est restrictive. h ne doit pas tendre vite vers 0 et k ne doit pas tendre rapidement vers l'infini. Par exemple, si on prend $h = 1/(\log n)^\tau$ alors $k \ll \log n / (\tau \log \log n)$. Dans les cas usuels, leur condition (H5b) est vérifiée dès que $(\log n)^2 = O(n\psi(h))$. Cette dernière condition est moins restrictive que $nh^k \rightarrow +\infty$, car k tend vers l'infini. Cependant, leur condition (H2) est très restrictive. Car s'ils prennent une semi-métrique à partir de projections sur un espace de dimension k fixe (comme leur exemple 4), la condition (H2) est difficile à vérifier.

6. Simulations

La première étude que nous avons effectuée consiste à estimer la fonction θ_0 par une méthode de Monte Carlo. On se fixe une base $\phi_j, j \geq 1$ qui est une base de Fourier $\phi_j(t) = \sqrt{2} \sin(\pi jt), t \in [0, 1], j \geq 1$. Nous définissons θ_0 dans cette base :

$$\theta_0(t) = \sum_{j=1}^{20} \theta_{0,j} \phi_j(t),$$

FIGURE 1. La fonction θ_0 .

où $\theta_{0,j} = 1/j$, $1 \leq j \leq 3$, $\theta_{0,j} = 0$ pour $j > 3$. La figure 1 représente la fonction θ_0 . Nous construisons ensuite un échantillon (x_i, y_i) pour $i = 1, \dots, 400$ comme suit

$$x(t) = \sum_{j=1}^{20} \varepsilon_j \phi_j(t),$$

où les ε_j sont générés par une loi normale centrée et de variance $1/j^2$. On choisit un lien logit avec $\rho(t) = \exp(t)/(1 + \exp(t))$ et les y_i sont générés par y qui suit une loi de Bernoulli de paramètre $p(x) = \rho(\sum_{j=1}^{20} \theta_{0,j} \varepsilon_j)$. Une fois obtenu l'échantillon, on estime θ_0 par la méthode décrite plus haut, en considérant les instruments optimaux connus, avec $p = 5$ et une base différente de ϕ_j à savoir une base spline à 5 éléments. La figure 2 représente les fonctions θ_0 et $\hat{\theta}$ obtenues en moyennant sur 1000 répliques la procédure.

7. Conclusion

Nous avons construit une procédure d'estimation dans une riche classe de modèles. Cependant, le modèle de régression quantile n'est pas inclus dans notre démarche. Une approche récente pour des données finies-dimensionnelles a été construite par [Lavergne and Patilea \(2013\)](#) pour palier entre autre à ce problème. Son extension aux données fonctionnelles fera l'objet d'un travail futur.

8. Preuves

Dans toutes les preuves, on prend $\hat{P} = I_p$, sans restriction de généralité.

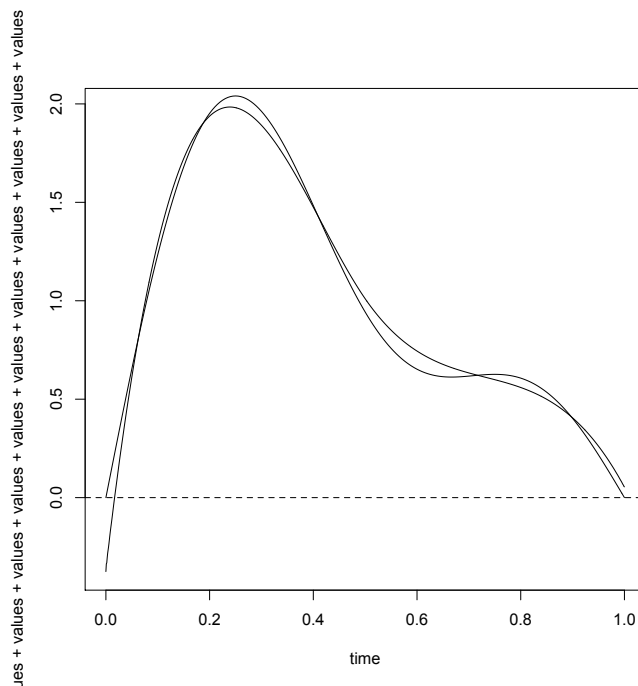


FIGURE 2. Estimation de θ_0 avec les instruments optimaux.

Preuve du lemme 4.1. On utilise une version du théorème 2.1 de [Newey and McFadden \(1994\)](#) :

Théorème 8.1. *On suppose qu'il existe une fonction aléatoire $\hat{Q}_n(\theta)$ telle que $\hat{\theta}$ maximise $\hat{Q}_n(\theta)$. On suppose de plus qu'il existe une fonction $Q_0(\theta)$ telle que :*

- (i) θ_0 est l'unique maximum de $Q_0(\theta)$,
- (ii) Θ est compact,
- (iii) $Q_0(\theta)$ est continue,
- (iv) on a

$$\sup_{\theta \in \Theta} |\hat{Q}_n(\theta) - Q_0(\theta)| \xrightarrow{\mathcal{P}} 0,$$

alors on obtient

$$\hat{\theta} \xrightarrow{\mathcal{P}} \theta_0.$$

Pour obtenir la convergence uniforme en probabilité de $\hat{Q}_n(\theta)$, on a besoin d'une loi uniforme des grands nombres, voir le lemme 2.4 de [Newey and McFadden \(1994\)](#) :

Lemme 8.2. *Sous les hypothèses suivantes :*

1. les données sont *i.i.d.*,
2. Θ est compact,
3. $a(z_i, \theta)$ est continue en chaque $\theta \in \Theta$ avec probabilité un,
4. Il existe $d(z)$ avec $\|a(z, \theta)\| \leq d(z) \forall \theta \in \Theta$ et $\mathbb{E}[d(z)] < \infty$,

on obtient la continuité de $\mathbb{E}[a(z, \theta)]$ et

$$\sup_{\theta \in \Theta} \left\| \frac{1}{n} \sum_{i=1}^n a(z_i, \theta) - \mathbb{E}[a(z, \theta)] \right\| \xrightarrow{\mathcal{P}} 0.$$

On pose $\hat{Q}_n(\theta) = \hat{g}'_n(\theta^{(p)}) \hat{g}_n(\theta^{(p)})$, avec $\hat{g}_n(\theta) = \frac{1}{n} \sum_{i=1}^n B(x_i^{(p)}) g(y_i, \langle x_i^{(p)}, \theta^{(p)} \rangle)$. Posons de plus

$$Q_0(\theta) = \lim_{n \rightarrow \infty} \left\| \mathbb{E} \left[B(x^{(p)}) g(y, \langle x^{(p)}, \theta^{(p)} \rangle) \right] \right\|^2,$$

on a $Q_0(\theta_0) = 0$ qui est l'unique minimum. En effet, s'il existe θ_1 tel que $Q_0(\theta_1) = 0$, alors en notant B_j la j -ème coordonnée de B , nous avons

$$\mathbb{E} \left[B_j(x^{(p)}) \mathbb{E}[g(y, \langle x^{(p)}, \theta_1^{(p)} \rangle | x^{(p)})] \right] \rightarrow 0.$$

D'où $\forall j$,

$$\mathbb{E} [B_j(x) \mathbb{E}[g(y, \langle x, \theta_1 \rangle | x)]] = 0.$$

D'où $\mathbb{E}[g(y, \langle x, \theta_1 \rangle | x)] = 0$ et donc $\theta_1 = \theta_0$.

On applique ensuite la loi uniforme des grands nombres, on a alors par l'hypothèse 4 qui est vérifiée par la condition (d) du lemme 4.1 :

$$\sup_{\theta \in \Theta} \left\| \hat{g}_n(\theta^{(p)}) - g_p(\theta^{(p)}) \right\| \xrightarrow{\mathcal{P}} 0,$$

où $g_p(\theta^{(p)}) = \mathbb{E}[B(x^{(p)}) g(y, \langle x^{(p)}, \theta^{(p)} \rangle)]$. On a

$$\begin{aligned} |\hat{Q}_n(\theta) - Q_0(\theta)| &\leq |\hat{Q}_n(\theta) - g'_p(\theta^{(p)}) g_p(\theta^{(p)})| + |g'_p(\theta^{(p)}) g_p(\theta^{(p)}) - Q_0(\theta)| \\ &\leq \|\hat{g}_n(\theta^{(p)}) - g_p(\theta^{(p)})\|^2 + 2\|g_p(\theta^{(p)})\| \|\hat{g}_n(\theta^{(p)}) - g_p(\theta^{(p)})\| \\ &\quad + |g'_p(\theta^{(p)}) g_p(\theta^{(p)}) - Q_0(\theta)|. \end{aligned}$$

D'où

$$\sup_{\theta \in \Theta} |\hat{Q}_n(\theta) - Q_0(\theta)| \xrightarrow[n \rightarrow \infty]{\mathcal{P}} 0.$$

Toutes les hypothèses du théorème sont vérifiées, on en déduit que

$$\|\hat{\theta}_n - \theta_0^{(p)}\| \xrightarrow[n \rightarrow \infty]{\mathcal{P}} 0.$$

□

Preuve du théorème 4.1. En posant

$$U(\theta) = \frac{\partial \hat{g}_n}{\partial \theta'}(\hat{\theta}) \hat{P} \hat{g}_n(\theta),$$

on a par Taylor, pour un $\bar{\theta}$ compris entre θ_0 et $\hat{\theta}$

$$0 = U(\hat{\theta}) = U(\theta_0) + \frac{\partial U}{\partial \theta}(\bar{\theta})(\hat{\theta} - \theta_0).$$

D'où

$$\sqrt{n}(\hat{\theta} - \theta_0) = - \left(\frac{\partial \hat{g}_n}{\partial \theta'}(\hat{\theta}) \hat{P} \frac{\partial \hat{g}_n}{\partial \theta}(\bar{\theta}) \right)^{-1} \sqrt{n}U(\theta_0),$$

qui se simplifie en

$$\sqrt{n}(\hat{\theta} - \theta_0) = - \left(\frac{\partial \hat{g}_n}{\partial \theta}(\bar{\theta}) \right)^{-1} \sqrt{n}\hat{g}_n(\theta_0).$$

Dans une première partie, on s'attache à remplacer $\bar{\theta}$ par θ_0 . $\sqrt{n}(\hat{\theta} - \theta_0)$ s'écrit

$$\begin{aligned} &= - \left[\frac{\partial \hat{g}_n}{\partial \theta}(\bar{\theta}) - \frac{\partial \hat{g}_n}{\partial \theta}(\theta_0) + \frac{\partial \hat{g}_n}{\partial \theta}(\theta_0) \right]^{-1} \sqrt{n}\hat{g}_n(\theta_0), \\ &= - \left\{ I_p + \left[\frac{\partial \hat{g}_n}{\partial \theta}(\theta_0) \right]^{-1} \left[\frac{\partial \hat{g}_n}{\partial \theta}(\bar{\theta}) - \frac{\partial \hat{g}_n}{\partial \theta}(\theta_0) \right] \right\}^{-1} \left[\frac{\partial \hat{g}_n}{\partial \theta}(\theta_0) \right]^{-1} \sqrt{n}\hat{g}_n(\theta_0). \end{aligned}$$

Comme pour toute matrice C avec $\|C\|_2 < 1$ on a $(I_p + C)^{-1} = I_p - C + R$ avec $\|R\|_2 \leq (1 - \|C\|_2)^{-1} \|C\|_2^2$, il suffit de montrer que

$$\left\| \left[\frac{\partial \hat{g}_n}{\partial \theta}(\theta_0) \right]^{-1} \right\|_2 = O_p(1) \text{ and } \left\| \frac{\partial \hat{g}_n}{\partial \theta}(\bar{\theta}) - \frac{\partial \hat{g}_n}{\partial \theta}(\theta_0) \right\|_2 = o_p(1). \quad (8.7)$$

Par la loi des grands nombres,

$$\frac{\partial \hat{g}_n}{\partial \theta}(\theta_0) - \mathbb{E} \left[\frac{\partial \hat{g}_n}{\partial \theta}(\theta_0) \right] = \frac{\partial \hat{g}_n}{\partial \theta}(\theta_0) - \Lambda^{-1} = o_p(1).$$

Ceci assure la première partie de l'équation (8.7). Pour la deuxième partie de l'équation (8.7), on peut écrire

$$\begin{aligned} \left\| \frac{\partial \hat{g}_n}{\partial \theta}(\bar{\theta}) - \frac{\partial \hat{g}_n}{\partial \theta}(\theta_0) \right\|_2 &= \frac{1}{n} \left\| \sum_{i=1}^n B(x_i) \left(\frac{\partial g}{\partial \theta}(z_i, \bar{\theta}) - \frac{\partial g}{\partial \theta}(z_i, \theta_0) \right) \right\|_2 \\ &\leq \frac{1}{n} \sum_{i=1}^n \|B(x_i)\| \left\| \frac{\partial g}{\partial \theta}(z_i, \bar{\theta}) - \frac{\partial g}{\partial \theta}(z_i, \theta_0) \right\|_2 \\ &\leq O_p(\|\bar{\theta} - \theta_0\|^a) \frac{1}{n} \sum_{i=1}^n \|B(x_i)\| |c(y_i)| \\ &= o_p(1), \end{aligned}$$

avec les hypothèses

$$\left\| \frac{\partial g}{\partial \theta}(z_i, b) - \frac{\partial g}{\partial \theta}(z_i, d) \right\|_2 \leq c(y_i) \|b - d\|^a$$

et

$$\|B_j(x_i)\| |c(y_i)| \leq \Gamma(z_i), \quad \mathbb{E}[\Gamma(z_i)] < \infty,$$

car

$$\left\| \frac{\partial g}{\partial \theta} \left(z_i, \sum_{j=1}^p \varepsilon_j^i \bar{\theta}_j \right) - \frac{\partial g}{\partial \theta} \left(z_i, \sum_{j=1}^p \varepsilon_j^i \theta_{0,j} \right) \right\|_2^2 \leq c(y_i) \left| \sum_{j=1}^p \varepsilon_j^i (\bar{\theta}_j - \theta_{0,j}) \right|^a$$

or par Cauchy-Schwarz, on a

$$\left| \sum_{j=1}^p \varepsilon_j^i (\bar{\theta}_j - \theta_{0,j}) \right| \leq \sqrt{\sum_{j=1}^{+\infty} \varepsilon_j^{i2}} \sqrt{\sum_{j=1}^p (\bar{\theta}_j - \theta_{0,j})^2} = O_p(1).$$

On étudie maintenant $Z_n' \Lambda^{-1} Z_n$ où $Z_n = \left(\frac{\partial \hat{g}_n}{\partial \theta}(\theta_0) \right)^{-1} \sqrt{n} \hat{g}_n(\theta_0)$. On cherche quel terme est dominant dans la décomposition suivante de $Z_n' \Lambda^{-1} Z_n$. En posant $\chi_n = \Lambda^{\frac{1}{2}} \sqrt{n} \hat{g}_n(\theta_0)$ et $\psi_n = \Lambda^{-\frac{1}{2}} \left(\frac{\partial \hat{g}_n}{\partial \theta}(\theta_0) \right)^{-1} \Lambda^{-\frac{1}{2}}$, on a alors la décomposition :

$$Z_n' \Lambda^{-1} Z_n = \chi_n' \chi_n + 2\chi_n' (\psi_n - I_p) \chi_n + \chi_n' (\psi_n - I_p) (\psi_n - I_p) \chi_n.$$

Il reste à montrer que $\|\psi_n - I_p\|_2^2 = O_p\left(\frac{1}{p}\right)$.

Étudions d'abord ψ_n^{-1} :

$$\begin{aligned} \psi_n^{-1} &= \Lambda^{-1/2} \frac{\partial \hat{g}_n}{\partial \theta}(\theta_0) \Lambda^{-1/2} \\ &= \left(\sum_{k_1, k_2=1}^p \xi_{l_1, k_1}^{(1/2)} \left(\frac{1}{n} \sum_{k=1}^n \left(\frac{\partial \tilde{g}_k}{\partial \theta} \right)_{k_1, k_2} \xi_{k_2, l_2}^{(1/2)} \right) \right)_{l_1, l_2=1}^p \\ &= \left(\frac{1}{n} \sum_{k_1, k_2=1}^p \xi_{l_1, k_1}^{(1/2)} \xi_{k_2, l_2}^{(1/2)} \sum_{k=1}^n \left(\frac{\partial \tilde{g}_k}{\partial \theta} \right)_{k_1, k_2} \right)_{l_1, l_2=1}^p \end{aligned}$$

d'où, le calcul suivant :

$$\begin{aligned} \mathbb{E}(\|\psi_n^{-1} - I_p\|_2^2) &= \mathbb{E} \left(\sum_{l_1, l_2=1}^p \left(\frac{1}{n} \sum_{k=1}^n \sum_{k_1, k_2=1}^p \xi_{l_1, k_1}^{(1/2)} \xi_{k_2, l_2}^{(1/2)} \left(\frac{\partial \tilde{g}_k}{\partial \theta} \right)_{k_1, k_2} - \delta_{l_1, l_2} \right) \right. \\ &\quad \times \left. \left(\frac{1}{n} \sum_{k'=1}^n \sum_{k'_1, k'_2=1}^p \xi_{l_1, k'_1}^{(1/2)} \xi_{k'_2, l_2}^{(1/2)} \left(\frac{\partial \tilde{g}'_{k'}}{\partial \theta} \right)_{k'_1, k'_2} - \delta_{l_1, l_2} \right) \right) \\ &= \mathbb{E} \left(\sum_{l_1, l_2=1}^p \left(\frac{1}{n^2} \sum_{k, k'=1}^n \sum_{k_1, k_2, k'_1, k'_2}^p \xi_{l_1, k_1}^{(1/2)} \xi_{k_2, l_2}^{(1/2)} \xi_{l_1, k'_1}^{(1/2)} \xi_{k'_2, l_2}^{(1/2)} \left(\frac{\partial \tilde{g}_k}{\partial \theta} \right)_{k_1, k_2} \left(\frac{\partial \tilde{g}'_{k'}}{\partial \theta} \right)_{k'_1, k'_2} \right. \right. \\ &\quad \left. \left. - 2\delta_{l_1, l_2} \frac{1}{n} \sum_{k=1}^n \sum_{k_1, k_2=1}^p \xi_{l_1, k_1}^{(1/2)} \xi_{k_2, l_2}^{(1/2)} \left(\frac{\partial \tilde{g}_k}{\partial \theta} \right)_{k_1, k_2} \right. \right. \\ &\quad \left. \left. + \delta_{l_1, l_2} \right) \right) \\ &= \sum_{l_1, l_2=1}^p \left(\frac{1}{n^2} \sum_{k, k'=1}^n \sum_{k_1, k_2, k'_1, k'_2}^p \xi_{l_1, k_1}^{(1/2)} \xi_{k_2, l_2}^{(1/2)} \xi_{l_1, k'_1}^{(1/2)} \xi_{k'_2, l_2}^{(1/2)} \mathbb{E} \left[\left(\frac{\partial \tilde{g}_k}{\partial \theta} \right)_{k_1, k_2} \left(\frac{\partial \tilde{g}'_{k'}}{\partial \theta} \right)_{k'_1, k'_2} \right] \right. \\ &\quad \left. - 2\delta_{l_1, l_2} \frac{1}{n} \sum_{k=1}^n \sum_{k_1, k_2=1}^p \xi_{l_1, k_1}^{(1/2)} \xi_{k_2, l_2}^{(1/2)} \Lambda_{k_1, k_2}^{-1} \right. \\ &\quad \left. + \delta_{l_1, l_2} \right) \end{aligned}$$

car $\mathbb{E}[(\frac{\partial \tilde{g}_k}{\partial \theta})_{k_1, k_2}] = \Lambda_{k_1, k_2}^{-1}$
 ainsi,

$$-2\delta_{l_1, l_2} \frac{1}{n} \sum_{k=1}^n \sum_{k_1, k_2=1}^p \xi_{l_1, k_1}^{(1/2)} \xi_{k_2, l_2}^{(1/2)} \Lambda_{k_1, k_2}^{-1} = -2\delta_{l_1, l_2} \sum_{k_1, k_2=1}^p \xi_{l_1, k_1}^{(1/2)} \xi_{k_2, l_2}^{(1/2)} \Lambda_{k_1, k_2}^{-1} = -2\delta_{l_1, l_2}^2 = -2\delta_{l_1, l_2}$$

d'où en reprenant le calcul de $\mathbb{E}(\|\psi_n^{-1} - I_p\|_2^2)$, et en scindant en deux termes suivant $k = k'$ ou $k \neq k'$, c-à-d suivant que les variables sont indépendantes ou non, on obtient :

$$\begin{aligned} & \mathbb{E}(\|\psi_n^{-1} - I_p\|_2^2) \\ &= \sum_{l_1, l_2=1}^p \left(\frac{1}{n^2} \sum_{k=1}^n \sum_{k_1, k_2, k'_1, k'_2}^p \xi_{l_1, k_1}^{(1/2)} \xi_{k_2, l_2}^{(1/2)} \xi_{l_1, k'_1}^{(1/2)} \xi_{k'_2, l_2}^{(1/2)} \mathbb{E}[(\frac{\partial \tilde{g}_k}{\partial \theta})_{k_1, k_2} (\frac{\partial \tilde{g}_k}{\partial \theta})_{k'_1, k'_2}] \right) \\ &+ \frac{1}{n^2} \sum_{k \neq k'}^n \sum_{k_1, k_2, k'_1, k'_2}^p \xi_{l_1, k_1}^{(1/2)} \xi_{k_2, l_2}^{(1/2)} \xi_{l_1, k'_1}^{(1/2)} \xi_{k'_2, l_2}^{(1/2)} \Lambda_{k_1, k_2}^{-1} \Lambda_{k'_1, k'_2}^{-1} - \delta_{l_1, l_2} \\ &= \sum_{l_1, l_2=1}^p \frac{1}{n^2} \sum_{k=1}^n \sum_{k_1, k_2, k'_1, k'_2}^p \xi_{l_1, k_1}^{(1/2)} \xi_{k_2, l_2}^{(1/2)} \xi_{l_1, k'_1}^{(1/2)} \xi_{k'_2, l_2}^{(1/2)} \mathbb{E}[(\frac{\partial \tilde{g}_k}{\partial \theta})_{k_1, k_2} (\frac{\partial \tilde{g}_k}{\partial \theta})_{k'_1, k'_2}] - \frac{1}{n} \delta_{l_1, l_2} \\ &= \frac{1}{n^2} \sum_{l_1, l_2=1}^p \sum_{k=1}^n \sum_{k_1, k_2, k'_1, k'_2}^p \xi_{l_1, k_1}^{(1/2)} \xi_{k_2, l_2}^{(1/2)} \xi_{l_1, k'_1}^{(1/2)} \xi_{k'_2, l_2}^{(1/2)} \mathbb{E}[(\frac{\partial \tilde{g}_k}{\partial \theta})_{k_1, k_2} (\frac{\partial \tilde{g}_k}{\partial \theta})_{k'_1, k'_2}] + O(\frac{p}{n}) \\ &= \frac{1}{n} \sum_{k_1, k_2, k'_1, k'_2}^p \xi_{k_1, k'_1} \xi_{k_2, k'_2} \mathbb{E}[(\frac{\partial \tilde{g}_k}{\partial \theta})_{k_1, k_2} (\frac{\partial \tilde{g}_k}{\partial \theta})_{k'_1, k'_2}] + O(\frac{p}{n}) \end{aligned}$$

par hypothèses $\sum_{k_1, k_2, k'_1, k'_2}^p \xi_{k_1, k'_1} \xi_{k_2, k'_2} \mathbb{E}[(\frac{\partial \tilde{g}_k}{\partial \theta})_{k_1, k_2} (\frac{\partial \tilde{g}_k}{\partial \theta})_{k'_1, k'_2}] = o(n/p^2)$. On a donc le résultat suivant $\|\psi_n^{-1} - I_p\|_2^2 = O_p(1/p^2)$. On utilise ensuite le résultat suivant :

$$\|\psi_n - I_p\|_2 \leq \|\psi_n\|_2 \|\psi_n^{-1} - I_p\|_2$$

conjugué avec ce résultat

$$\|\psi_n\|_2 \leq \frac{\|I_p\|_2}{1 - \|\psi_n^{-1} - I_p\|_2},$$

cela nous donne le résultat escompté, à savoir $\|\psi_n - I_p\|_2 = O_p(\frac{1}{p})$.

Traisons le terme dominant : On va montrer que

$$\frac{\chi_n' \chi_n - p}{\sqrt{2p}} \longrightarrow N(0, 1).$$

Pour ce faire, on décompose en deux termes :

$$\begin{aligned} \chi_n' \chi_n &= n \hat{g}_n(\theta_0)' \Lambda \hat{g}_n(\theta_0) \\ &= \frac{1}{n} \sum_{i=1}^n (B(x_i)g(z_i, \theta_0))' \Lambda (B(x_i)g(z_i, \theta_0)) \\ &+ \frac{1}{n} \sum_{i \neq k}^n (B(x_i)g(z_i, \theta_0))' \Lambda (B(x_k)g(z_k, \theta_0)) \\ &= A_n + B_n \end{aligned}$$

on a $\mathbb{E}[A_n] = p$ ainsi que $\text{Var}(A_n) = o(p)$ ce qui nous permet d'écrire

$$\frac{A_n - p}{\sqrt{p}} \rightarrow 0$$

En effet,

$$\begin{aligned} \mathbb{E}[A_n] &= \mathbb{E}[g'(z_1, \theta_0)B'(x_1)\Lambda B(x_1)g(z_1, \theta_0)] \\ &= \mathbb{E}[\text{tr}(g'B'\Lambda Bg)] \\ &= \mathbb{E}[\text{tr}(g'B'\Lambda^{\frac{1}{2}}\Lambda^{\frac{1}{2}}Bg)] \\ &= \mathbb{E}[\text{tr}(\Lambda^{\frac{1}{2}}Bgg'B'\Lambda^{\frac{1}{2}})] \\ &= \text{tr}(\Lambda^{\frac{1}{2}}\mathbb{E}[Bgg'B']\Lambda^{\frac{1}{2}}) \end{aligned}$$

ensuite on a $\mathbb{E}[Bgg'B'] = \mathbb{E}[\mathbb{E}[Bgg'B'|x_1]]$ et donc $\mathbb{E}[Bgg'B'] = \Lambda^{-1}$

$$\begin{aligned} \mathbb{E}[A_n] &= \text{tr}(\Lambda^{\frac{1}{2}}\Lambda^{-1}\Lambda^{\frac{1}{2}}) \\ &= \text{tr}(I_p) \\ &= p \end{aligned}$$

On calcule maintenant la variance : pour cela calculons $\mathbb{E}[A_n^2]$

$$A_n^2 = \frac{1}{n^2} \sum_{i,j=1}^n \tilde{g}'^{(i)} \Lambda \tilde{g}^{(i)} \tilde{g}'^{(j)} \Lambda \tilde{g}^{(j)} \quad (8.8)$$

$$\begin{aligned} \mathbb{E}[A_n^2] &= \frac{1}{n^2} \sum_{i,j=1}^n \mathbb{E}(\tilde{g}'^{(i)} \Lambda \tilde{g}^{(i)} \tilde{g}'^{(j)} \Lambda \tilde{g}^{(j)}) \\ &= \frac{1}{n^2} \sum_{i=1}^n \mathbb{E}(\tilde{g}'^{(i)} \Lambda \tilde{g}^{(i)} \tilde{g}'^{(i)} \Lambda \tilde{g}^{(i)}) \\ &\quad + \frac{1}{n^2} \sum_{i \neq j}^n \mathbb{E}(\tilde{g}'^{(i)} \Lambda \tilde{g}^{(i)} \tilde{g}'^{(j)} \Lambda \tilde{g}^{(j)}) \end{aligned}$$

Or $\mathbb{E}(\tilde{g}'^{(i)} \Lambda \tilde{g}^{(i)} \tilde{g}'^{(j)} \Lambda \tilde{g}^{(j)}) = (\mathbb{E}(\tilde{g}'^{(i)} \Lambda \tilde{g}^{(i)}))^2 = p^2$

$$\begin{aligned} \mathbb{E}[A_n^2] &= \frac{n(n-1)}{n^2} p^2 + \frac{1}{n^2} \sum_{i=1}^n \mathbb{E}(\tilde{g}'^{(i)} \Lambda \tilde{g}^{(i)} \tilde{g}'^{(i)} \Lambda \tilde{g}^{(i)}) \\ &= p^2 - \frac{p^2}{n} + \frac{1}{n} \mathbb{E}(\tilde{g}'^{(1)} \Lambda \tilde{g}^{(1)} \tilde{g}'^{(1)} \Lambda \tilde{g}^{(1)}) \end{aligned}$$

On utilise alors l'hypothèse D-(e), en effet

$$\mathbb{E}(\tilde{g}'^{(1)} \Lambda \tilde{g}^{(1)} \tilde{g}'^{(1)} \Lambda \tilde{g}^{(1)}) = \sum_{k_1, k_2, k_3, k_4=1}^p \mathbb{E}[\tilde{g}_{k_1} \tilde{g}_{k_2} \tilde{g}_{k_3} \tilde{g}_{k_4}] \xi_{k_1, k_2} \xi_{k_3, k_4} = o(n/p^2) \quad (8.9)$$

On obtient donc $\text{Var}(A_n) = o(p)$ d'où le résultat.

En utilisant un T.C.L pour martingales démontré par [Brown \(1971\)](#), on montre que $B_n/\sqrt{2p} \rightarrow N(0, 1)$. Rappelons ce que vaut B_n :

$$B_n = \frac{1}{n} \sum_{i \neq k}^n (B(x_i)g(z_i, \theta_0))' \Lambda(B(x_k)g(z_k, \theta_0)),$$

Il faut centrer pour pouvoir appliquer le théorème. Pour cela, posons

$$\widetilde{B}_n = \frac{1}{n} \sum_{i \neq k} \widetilde{g}_c^{(i)'} \Lambda \widetilde{g}_c^{(k)},$$

où $\widetilde{g}_c^{(j)} = B(x_j)g(z_j, \theta_0) - \mathbb{E}[B(x_1)g(z_1, \theta_0)]$. On obtient alors la décomposition de B_n en quatre termes :

$$\begin{aligned} B_n = \widetilde{B}_n &+ \frac{n-1}{n} (\mathbb{E}B(x_1)g(z_1, \theta_0))' \Lambda \left(\sum_{k=1}^n B(x_k)g(z_k, \theta_0) \right) \\ &+ \frac{n-1}{n} \left(\sum_{k=1}^n B(x_k)g(z_k, \theta_0) \right)' \Lambda (\mathbb{E}B(x_1)g(z_1, \theta_0)) \\ &- (n-1) (\mathbb{E}B(x_1)g(z_1, \theta_0))' \Lambda (\mathbb{E}B(x_1)g(z_1, \theta_0)) \end{aligned}$$

Remarquons que :

$$\mathbb{E}[B(x_1)g(z_1, \theta_0)] = O(pr_n).$$

Ainsi le dernier terme est de l'ordre $O(np_n^2 r_n^2)$, le second et le troisième terme sont de l'ordre $O_{\mathbb{P}}(np_n^2 r_n^2)$. D'où en divisant par $\sqrt{2p}$ et en utilisant l'hypothèse $n^{11/8} r_n^2 \rightarrow 0$, ces termes sont négligeables par rapport à \widetilde{B}_n . Etudions maintenant \widetilde{B}_n . Posons

$$W_{nj} = \sum_{k=1}^{j-1} \widetilde{g}_c^{(j)'} \Lambda \widetilde{g}_c^{(k)}$$

on obtient alors $\widetilde{B}_n = \frac{2}{n} \sum_{j=1}^n W_{nj}$.

Les $\{W_{nj}, 1 \leq j \leq n, n \in \mathbb{N}\}$ forment un tableau triangulaire de différences de martingales associées aux filtrations $\mathcal{F}_{nj} = \sigma(\varepsilon_t^{(i)}, y_i, 1 \leq i \leq j, 1 \leq t \leq p) (1 \leq j \leq n, n \in \mathbb{N})$.

En effet, $W_{nj} \in \mathcal{F}_{nj}$ et

$$\mathbb{E}[W_{nj} | \mathcal{F}_{n,j-1}] = 0,$$

car $\mathbb{E}[\widetilde{g}_c^{(j)} | \mathcal{F}_{n,j-1}] = 0$. Pour appliquer le théorème de [Brown \(1971\)](#), il faut valider la condition de normalisation conditionnelle ainsi que la condition de Lyapunov conditionnelle. En écrivant, $\widetilde{W}_{nj} = \frac{2}{n\sqrt{2p}} W_{nj}$, la condition de normalisation conditionnelle s'écrit :

$$\sum_{j=1}^n \mathbb{E}[\widetilde{W}_{nj}^2 | \mathcal{F}_{n,j-1}] \xrightarrow{p} 1,$$

la condition de Lyapunov conditionnelle s'écrit :

$$\sum_{j=1}^n \mathbb{E}[\widetilde{W}_{nj}^4 | \mathcal{F}_{n,j-1}] \xrightarrow{P} 0.$$

Pour cela calculons W_{nj}^2 :

$$\begin{aligned} W_{nj}^2 &= \sum_{k,k'=1}^{j-1} \tilde{g}_c^{(j)} \Lambda \tilde{g}_c^{(k)} \tilde{g}_c^{(j)} \Lambda \tilde{g}_c^{(k')} \\ &= \sum_{k,k'=1}^{j-1} \sum_{t_1, t_2, t_3, t_4}^p \tilde{g}_{c,t_1}^{(j)} \xi_{t_1, t_2} \tilde{g}_{c,t_2}^{(k)} \tilde{g}_{c,t_3}^{(j)} \xi_{t_3, t_4} \tilde{g}_{c,t_4}^{(k')} \\ &= \sum_{k,k'=1}^{j-1} \sum_{t_1, t_2, t_3, t_4}^p \tilde{g}_{c,t_1}^{(j)} \tilde{g}_{c,t_3}^{(j)} \tilde{g}_{c,t_2}^{(k)} \tilde{g}_{c,t_4}^{(k')} \xi_{t_1, t_2} \xi_{t_3, t_4}, \end{aligned}$$

d'où, en prenant l'espérance conditionnelle,

$$\mathbb{E}[W_{nj}^2 | \mathcal{F}_{n,j-1}] = \sum_{k,k'=1}^{j-1} \sum_{t_1, t_2, t_3, t_4}^p \mathbb{E}[\tilde{g}_{c,t_1}^{(j)} \tilde{g}_{c,t_3}^{(j)} \tilde{g}_{c,t_2}^{(k)} \tilde{g}_{c,t_4}^{(k')} \xi_{t_1, t_2} \xi_{t_3, t_4}].$$

Or,

$$\mathbb{E}[\tilde{g}_{c,t_1}^{(j)} \tilde{g}_{c,t_3}^{(j)}] = (\Lambda^{-1})_{t_1, t_3} - \mathbb{E}[\tilde{g}_{t_1}^{(j)}] \mathbb{E}[\tilde{g}_{t_3}^{(j)}],$$

donc :

$$\begin{aligned} \mathbb{E}[W_{nj}^2 | \mathcal{F}_{n,j-1}] &= \sum_{k,k'=1}^{j-1} \sum_{t_2, t_4}^p \mathbb{E}[\tilde{g}_{c,t_2}^{(k)} \tilde{g}_{c,t_4}^{(k')} \xi_{t_2, t_4}] \\ &\quad - \sum_{k,k'=1}^{j-1} \sum_{t_1, t_2, t_3, t_4}^p \mathbb{E}[\tilde{g}_{t_1}^{(j)}] \mathbb{E}[\tilde{g}_{t_3}^{(j)}] \mathbb{E}[\tilde{g}_{c,t_2}^{(k)} \tilde{g}_{c,t_4}^{(k')} \xi_{t_1, t_2} \xi_{t_3, t_4}]. \end{aligned}$$

On peut donc calculer son espérance :

$$\begin{aligned} \mathbb{E}[\mathbb{E}[W_{nj}^2 | \mathcal{F}_{n,j-1}]] &= \sum_{k=1}^{j-1} \sum_{t_2, t_4}^p \mathbb{E}[\tilde{g}_{c,t_2}^{(k)} \tilde{g}_{c,t_4}^{(k)} \xi_{t_2, t_4}] \\ &\quad - \sum_{k=1}^{j-1} \sum_{t_1, t_2, t_3, t_4}^p \mathbb{E}[\tilde{g}_{t_1}^{(j)}] \mathbb{E}[\tilde{g}_{t_3}^{(j)}] \mathbb{E}[\tilde{g}_{c,t_2}^{(k)} \tilde{g}_{c,t_4}^{(k)} \xi_{t_1, t_2} \xi_{t_3, t_4}] \\ &= (j-1)p - (j-1) \sum_{t_2, t_4}^p \mathbb{E}[\tilde{g}_{t_2}^{(k)}] \mathbb{E}[\tilde{g}_{t_4}^{(k)}] \xi_{t_2, t_4} \\ &\quad - (j-1) \sum_{t_1, t_2, t_3, t_4}^p \mathbb{E}[\tilde{g}_{t_1}^{(1)}] \mathbb{E}[\tilde{g}_{t_3}^{(1)}] \mathbb{E}[\tilde{g}_{c,t_2}^{(1)} \tilde{g}_{c,t_4}^{(1)} \xi_{t_1, t_2} \xi_{t_3, t_4}] \end{aligned}$$

maintenant,

$$\begin{aligned}
 & \mathbb{E}\left[\sum_{j=1}^n \mathbb{E}[\widetilde{W}_{nj}^2 | \mathcal{F}_{n,j-1}]\right] \\
 = & \sum_{j=1}^n \mathbb{E}\left[\frac{2}{n^2 p} W_{nj}^2 | \mathcal{F}_{n,j-1}\right] \\
 = & \frac{2}{n^2 p} \sum_{j=1}^n (j-1) \\
 & \times \left(p - \sum_{t_2, t_4}^p \mathbb{E}[\tilde{g}_{t_2}^{(k)}] \mathbb{E}[\tilde{g}_{t_4}^{(k)}] \xi_{t_2, t_4} - \sum_{t_1, t_2, t_3, t_4}^p \mathbb{E}[\tilde{g}_{t_1}^{(1)}] \mathbb{E}[\tilde{g}_{t_3}^{(1)}] \mathbb{E}[\tilde{g}_{c, t_2}^{(1)} \tilde{g}_{c, t_4}^{(1)}] \xi_{t_1, t_2} \xi_{t_3, t_4} \right) \\
 = & \frac{n(n-1)}{n^2} \left(1 - \frac{1}{p} O(p_n^2 r_n^2) - \frac{1}{p} O(p_n^4 r_n^4) \right) \\
 \rightarrow & 1
 \end{aligned}$$

Nous pouvons maintenant calculer la variance. Pour cela calculons, pour $k < j$:

$$\mathbb{E}[\mathbb{E}[W_{nj}^2 | \mathcal{F}_{n,j-1}] \mathbb{E}[W_{nk}^2 | \mathcal{F}_{n,k-1}]].$$

En notant

$$\begin{aligned}
 \mathbb{E}[W_{nj}^2 | \mathcal{F}_{n,j-1}] &= \sum_{k, k'=1}^{j-1} \sum_{t_2, t_4}^p \tilde{g}_{c, t_2}^{(k)} \tilde{g}_{c, t_4}^{(k')} \xi_{t_2, t_4} \\
 &- \sum_{k, k'=1}^{j-1} \sum_{t_1, t_2, t_3, t_4}^p \mathbb{E}[\tilde{g}_{t_1}^{(j)}] \mathbb{E}[\tilde{g}_{t_3}^{(j)}] \tilde{g}_{c, t_2}^{(k)} \tilde{g}_{c, t_4}^{(k')} \xi_{t_1, t_2} \xi_{t_3, t_4} \\
 &= M_j - N_j,
 \end{aligned}$$

on peut écrire

$$\begin{aligned}
 \mathbb{E}[\mathbb{E}[W_{nj}^2 | \mathcal{F}_{n,j-1}] \mathbb{E}[W_{nk}^2 | \mathcal{F}_{n,k-1}]] &= \mathbb{E}[(M_j - N_j)(M_k - N_k)] \\
 &= \mathbb{E}[M_j M_k - M_j N_k - M_k N_j + N_j N_k],
 \end{aligned}$$

où le terme dominant de cette somme est $\mathbb{E}[M_j M_k]$. Les trois autres termes sont négligeables par la condition D-(b). On calcule $\mathbb{E}[M_j M_k]$:

$$\begin{aligned}
 \mathbb{E}[M_j M_k] &= \sum_{i_1, i_2, i_3, i_4}^{j-1, k-1} \sum_{t_1, t_2, t_3, t_4}^p \mathbb{E}[\tilde{g}_{c, t_1}^{(i_1)} \tilde{g}_{c, t_2}^{(i_2)} \tilde{g}_{c, t_3}^{(i_3)} \tilde{g}_{c, t_4}^{(i_4)}] \xi_{t_1, t_2} \xi_{t_3, t_4} \\
 &= (k-1) \sum_{t_1, t_2, t_3, t_4}^p \mathbb{E}[\tilde{g}_{c, t_1} \tilde{g}_{c, t_2} \tilde{g}_{c, t_3} \tilde{g}_{c, t_4}] \xi_{t_1, t_2} \xi_{t_3, t_4} \\
 &+ (j-1)(k-1) \sum_{t_1, t_2, t_3, t_4}^p \mathbb{E}[\tilde{g}_{c, t_1} \tilde{g}_{c, t_2}] \mathbb{E}[\tilde{g}_{c, t_3} \tilde{g}_{c, t_4}] \xi_{t_1, t_2} \xi_{t_3, t_4} \\
 &+ 2(k-1)^2 \sum_{t_1, t_2, t_3, t_4}^p \mathbb{E}[\tilde{g}_{c, t_1} \tilde{g}_{c, t_3}] \mathbb{E}[\tilde{g}_{c, t_2} \tilde{g}_{c, t_4}] \xi_{t_1, t_2} \xi_{t_3, t_4}
 \end{aligned}$$

or on a :

$$\sum_{t_1, t_2, t_3, t_4}^p \mathbb{E}[\tilde{g}_{c, t_1} \tilde{g}_{c, t_2}] \mathbb{E}[\tilde{g}_{c, t_3} \tilde{g}_{c, t_4}] \xi_{t_1, t_2} \xi_{t_3, t_4} = p^2 - 2pO(p^2 r_n^2) + O(p^4 r_n^4).$$

On peut aussi remplacer $\mathbb{E}[\tilde{g}_{c, t_1} \tilde{g}_{c, t_2} \tilde{g}_{c, t_3} \tilde{g}_{c, t_4}]$ par $\mathbb{E}[\tilde{g}_{t_1} \tilde{g}_{t_2} \tilde{g}_{t_3} \tilde{g}_{t_4}]$ en remarquant que c'est le terme dominant parmi la décomposition en seize termes. On obtient donc

$$\begin{aligned} & \mathbb{E}[(\sum_{j=1}^n \mathbb{E}[W_{nj}^2 | \mathcal{F}_{n, j-1}])^2] \\ &= \sum_{j=1}^n \mathbb{E}[(\mathbb{E}[W_{nj}^2 | \mathcal{F}_{n, j-1}])^2] + 2 \sum_{k < j} \mathbb{E}[\mathbb{E}[W_{nj}^2 | \mathcal{F}_{n, j-1}] \mathbb{E}[W_{nk}^2 | \mathcal{F}_{n, k-1}]] \\ &= \sum_{j=1}^n (j-1) \sum_{t_1, t_2, t_3, t_4}^p \mathbb{E}[\tilde{g}_{t_1} \tilde{g}_{t_2} \tilde{g}_{t_3} \tilde{g}_{t_4}] \xi_{t_1, t_2} \xi_{t_3, t_4} + (j-1)^2 p^2 + 2(j-1)^2 p \\ &+ 2 \sum_{j=1}^n \sum_{k=1}^{j-1} ((k-1) \sum_{t_1, t_2, t_3, t_4}^p \mathbb{E}[\tilde{g}_{t_1} \tilde{g}_{t_2} \tilde{g}_{t_3} \tilde{g}_{t_4}] \xi_{t_1, t_2} \xi_{t_3, t_4} + (j-1)(k-1)p^2 + 2(k-1)^2 p) + o(1) \\ &= O(n^3 \sum_{t_1, t_2, t_3, t_4}^p \mathbb{E}[\tilde{g}_{t_1} \tilde{g}_{t_2} \tilde{g}_{t_3} \tilde{g}_{t_4}] \xi_{t_1, t_2} \xi_{t_3, t_4}) + \frac{n^4}{4} p^2 (1 + o(1)) + \frac{n^4}{6} p (1 + o(1)) \end{aligned}$$

On trouve que $\mathbb{E}[(\sum_{j=1}^n \mathbb{E}[\tilde{W}_{nj}^2 | \mathcal{F}_{n, j-1}])^2] = 1 + o(1)$, et donc $\text{Var}(\sum_{j=1}^n \mathbb{E}[\tilde{W}_{nj}^2 | \mathcal{F}_{n, j-1}]) \rightarrow 0$

La condition de Lyapunov conditionnelle se traite de la même façon en utilisant les hypothèses D-(b) et D-(f). \square

Preuve du théorème 5.1.

Lemme 8.3. *On a*

$$\begin{aligned} C' \psi(h) &\leq \mathbb{E}K(h^{-1} \|X_1 - x\|) \leq C \psi(h), \\ C_1 \psi(h) &\leq \mathbb{E}K^2(h^{-1} \|X_1 - x\|) \leq C_2 \psi(h) \end{aligned}$$

et

$$\text{Var} \left(\frac{1}{n} \sum_{i=1}^n K(h^{-1} \|X_i - x\|) \right) \rightarrow 0$$

Démonstration. La preuve est similaire au lemme 4.4 de [Ferraty and Vieu \(2006\)](#). Les hypothèses sur le noyau K et sur les petites boules de probabilité permettent de dire que le noyau est un noyau de type II et la condition additionnelle est vérifiée ; ce qui permet d'appliquer le lemme 4.4. Concernant la variance, on peut écrire où C et C' sont des constantes :

$$\begin{aligned} \text{Var} \left(\frac{1}{n} \sum_{i=1}^n K(h^{-1} \|X_i - x\|) \right) &= \frac{1}{n} \text{Var} (K(h^{-1} \|X_k - x\|)) \\ &\leq \frac{1}{n} (C \psi(h) - C' \psi^2(h)) \\ &\rightarrow 0 \end{aligned}$$

On a utilisé le fait que $\psi(h) \rightarrow 0$ quand h tend vers 0, ainsi que les deux premières propriétés. \square

Lemme 8.4.

$$\sup_{x \in S} |\hat{m}(x, t) - \hat{m}(x^{(k)}, t)| = o_{\mathbb{P}}(1).$$

Démonstration. Pour cela, on utilise le caractère Lipchitzien de K . On peut donc écrire :

$$\begin{aligned} |K(h^{-1}\|X_i - x\|) - K(h^{-1}\|X_i - x^{(k)}\|)| &\leq \frac{C}{h} \left| \|X_i - x\| - \|X_i - x^{(k)}\| \right| \\ &\leq \frac{C}{h} \|x - x^{(k)}\| \\ &\leq \frac{C}{h} \sup_{x \in S} \|x - x^{(k)}\|. \end{aligned}$$

Donc,

$$\begin{aligned} &|\hat{m}(x, t) - \hat{m}(x^{(k)}, t)| \\ &= \frac{1}{n} \left| \sum_{j=1}^n \Phi(Y_j, t) \left(\frac{K(h^{-1}\|X_j - x\|)}{\frac{1}{n} \sum_{i=1}^n K(h^{-1}\|X_i - x\|)} - \frac{K(h^{-1}\|X_j - x^{(k)}\|)}{\frac{1}{n} \sum_{i=1}^n K(h^{-1}\|X_i - x^{(k)}\|)} \right) \right| \\ &\leq \frac{C}{h\psi(h)} \|x - x^{(k)}\| O_{\mathbb{P}}(1). \end{aligned}$$

D'où le résultat. □

Lemme 8.5.

$$\sup_{x \in S, t \in \mathbb{R}} |m(x^{(k)}, t) - m(x, t)| = o(1)$$

Démonstration. En effet,

$$\begin{aligned} |m(x^{(k)}, t) - m(x, t)| &\leq C(t) \|x^{(k)} - x\|_{L^2} \\ &\leq \sup_{t \in \mathbb{R}} C(t) \sup_{x \in S} \|x^{(k)} - x\|_{L^2}. \end{aligned}$$

D'où le résultat, en faisant tendre k vers l'infini. □

Lemme 8.6.

$$\sup_{x \in S, t \in \mathbb{R}} |\hat{m}(x^{(k)}, t) - m(x^{(k)}, t)| = o_{\mathbb{P}}(1)$$

Démonstration. On applique le théorème 2 voir aussi le corollaire 2 d'[Einmahl and Mason \(2005\)](#). □

Au vu des différents lemmes, nous obtenons le résultat. □

Références

- Brown, B. M. (1971). Martingale central limit theorems. *Ann. Math. Statist.*, 42 :59–66.
 Chen, D., Hall, P., and Müller, H.-G. (2011). Single and multiple index functional regression models with nonparametric link. *The Annals of Statistics*, 39(3) :1720–1747.
 Domínguez, M. A. and Lobato, I. N. (2004). Consistent estimation of models defined by conditional moment restrictions. *Econometrica*, 72(5) :1601–1615.

- Einmahl, U. and Mason, D. M. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *Ann. Statist.*, 33(3) :1380–1403.
- Ferraty, F., Laksaci, A., Tadj, A., and Vieu, P. (2010). Rate of uniform consistency for nonparametric estimates with functional variables. *J. Statist. Plann. Inference*, 140(2) :335–352.
- Ferraty, F. and Vieu, P. (2006). *Nonparametric functional data analysis*. Springer Series in Statistics. Springer, New York. Theory and practice.
- Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4) :1029–1054.
- Kitamura, Y., Tripathi, G., and Ahn, H. (2004). Empirical likelihood-based inference in conditional moment restriction models. *Econometrica*, 72(6) :1667–1714.
- Lavergne, P. and Patilea, V. (2013). Smooth minimum distance estimation and testing in conditional moment restrictions models : Uniform in bandwidth theory. *accept in Journal of econometrics*.
- Müller, H.-G. and Stadtmüller, U. (2005). Generalized functional linear models. *Ann. Statist.*, 33(2) :774–805.
- Newey, W. K. (1993). Efficient estimation of models with conditional moment restrictions. In *Econometrics*, volume 11 of *Handbook of Statist.*, pages 419–454. North-Holland, Amsterdam.
- Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. In *Handbook of econometrics*, Vol. IV, volume 2 of *Handbooks in Econom.*, pages 2111–2245. North-Holland, Amsterdam.
- Smith, R. J. (2007). Efficient information theoretic inference for conditional moment restrictions. *J. Econometrics*, 138(2) :430–460.