République Algérienne Démocratique et Populaire

Ministère de l'enseignement supérieure et de la recherche scientifique

N° Attribué par la bibliothèque

Année univ.: 2017/2018

Some Asymptotics Properties of Semiparametric Estimation.

Mémoire présenté en vue de l'obtention du diplôme de

Master Académique

Université de Saida - Dr Moulay Tahar

Discipline : MATHEMATIQUES

Spécialité : Analyse Stochastique, Statistique des Processus et Applications (ASSPA)

par

Mokhtaria BELOUFA [1]

Sous la direction de

Dr. Fethi MADANI

Soutenue le 20/06/2018 devant le jury composé de

M. LAOUNI	Université Dr Tahar Moulay - Saïda	Président
F. MADANI	Université Dr Tahar Moulay - Saïda	Encadreur
F. MOKHTARI	Université Dr Tahar Moulay - Saïda	Examinatrice
F. BENZIADI	Université Dr Tahar Moulay - Saïda	Examinatrice

1. e-mail : beloufa258@gmail.com

Table des matières

Remerciement

En préambule à ce mémoire nous remerciant ALLAH qui nous aide et nous donne la patience et le courage durant ces langues années d'étude.

Je tiens à exprimer toute ma reconnaissance à mon Directeur de mémoire monsieur : MADANI Fethi. Je la remercie de m'avoir encadré, orienté, aidé et conseillé.

Je remercie mes très chers parents, qui ont toujours été là pour moi.

J'adresse mes sincères remerciements à tous les professeurs, intervenants et toutes les personnes qui par leurs paroles, leurs écrits, leurs conseils et leurs critiques ont guidé mes réflexions et ont accepté à me rencontrer et répondre à mes questions durant la période de mes études.

Je remercie les membres du jury d'avoir pris la peine de lire et de juger ce travail.

Chaleureux remerciement à ma famille qui m'a soutenue moralement,

Je remercie très spécialement Mlle Fatima benziadi ,Mme Rouane Rachida,Mlle Idrissi Soumia, Mr Djelouli Mohamed , Mr Guendouzi Abdelhak et aussi Mr Jorfi Kouider.

Tous mes remerciements aux enseignants de département de mathématiques.

Merci à tous les collègues étudiantes et étudiants de Mathématiques .

Enfin, à tous les personnes qui me aidé lors de la réalisation de ce modeste travail.

"***Merci à tous***"

Dédicace

Au mon dieu le clément et le miséricordieux louange à ALLAH le tout puissant.

Je dédie ce travail :

À mes chers parents, ma famille et à tous ce qui me sont proches.

Résumé

Les modèles semi-paramétriques présente un compromis entre l'approche paramétrique (linéaire ou non linéaire) et l'approche non paramétrique. L'intervention de ces modèles nous assure une convergence plus rapide des estimateurs, ceci ne ramène à une réduction de la dimension, mais au prix d'un risque plus élevé d'erreur de spécification. Notons que nous distinguons à titre d'exemple sur les modèles de régression semi-paramétrique deux types de modèle : les régressions partiellement linéaires, les modèles d'indice simple. Dans ce travail nous présentons les modèles semi-paramétrique pour la regression ensuite nous étudions la normalité asymptotique de l'estimateur de la fonctions de distrubition dans un modèle à indice simple.

Introduction générale

En générale la littérature sur la modélisation semiparamétrique pour des donnée fonctionnelles est très restreinte. Disons que la première étude est due à Ferré et al.(2001)[18] qui ont proposé une extension de la méthode SIR (sliced inverse régression) aux variables fonctionnelles en utilisant des outils similaires à l'approche théorique de Cardot et al. (1999)[10]. Pour un modèle d'indice fonctionnel de la fonction de régression Ferraty et al.(2003)[17] ont obtenu des propriétés asymptotiques, lorsque l'indice est connu. Ils ont établi dans le cas i.i.d. la convergence prêsque complète de l'estimateur à noyau de la régression pour ce modèle. Leurs résultats ont été étendu au cas dépendent par Aït Saidi et al.(2005)[1].

Une étude générale est faite par Aït Saidi et al.(2008)[2], lorsque l'indice fonctionnel est inconnu. Ils ont proposé un estimateur de ce paramètre, basant sur la technique de validation croisée. Dabo-Niang et Serge (2010)[14], utilise un modèle semi-paramétrique fonctionnelle, où la variable réponse (à valeurs réelles) est expliquée par la somme d'une combinaison linéaire de composantes supposées inconnues d'une variable aléatoire multivariée et une transformation inconnu d'une variable aléatoire fonctionnelle.
Leur étude est concentrée sur l'estimation paramétrique des coefficients de la combinaison linéaire du modèle présenté, ils utilisent la méthode non paramétrique pour estimer la variable explicative fonctionnelle, basée sur l'approche des moindres carrés généralisés pour obtenir un estimateur de ces coefficients. Ils ont établi sous des hypothèses standards, la consistance forte ainsi que la normalité asymptotique de l'estimateur. Ils illustrent l'étude asymptotique de l'estimateur par des simulations du type Monte Carlo.

Dans ce qui suit, on présent les types de modèles statistique pour clotirer la partie introduction. Le premier chapitre est conscré à la présentation des modèles semi-paramétriques ainsi que des outils nécessaire pour ce type de modèles. Le dernier chapitre, on présente l'estimateur de la fonction de la distrubition dans un modèle à indice simple.

Modèles paramétriques :

Pour les modèles de régression paramétriques, la fonction de lien, entre la variable à expliquer $y \in \mathbb{R}$ et une variable explicative $x \in \mathbb{R}^p$, dépend d'un nombre fini de paramètres à estimer. Ils sont de la forme :

$$y = f_\theta(x) + \varepsilon$$

où f_θ appartient à une famille de fonctions paramétrées par θ, vecteur de paramètres réels, et où ε est un terme d'erreur aléatoire.
Dans un tel modéle, l'objectif est l'estimation du paramètre θ. Les techniques d'estimations paramétriques (maximum de vraisemblance, moindres carrés, ...) sont efficaces quand la famille de f_θ est correctement spécifiée. Ces modèles permettent une interprétation claire de l'impact de la variable explicative sur la variable à expliquer.
Cependant, le choix d'un bon modèle paramétrique au vu des données n'est pas toujours évident. Ainsi le modéle paramétrique choisi peut ne pas être en adéquation avec les données réelles et donc parfois être trés "éloigné" de la réalité des données. En conséquence les conclusions en découlant peuvent alors être erronées.
Afin de surmonter ce probléme de sélection du bon modèle paramétrique, les modèles non paramétriques ont été proposés.

Modèles non paramétriques :

Les modèles de régression non paramétriques apparaissent comme une alternative qui offre la fléxibilité désirée dans la modélisation (car aucune hypothèse paramétrique n'est imposée dans le modèle, seules des hypothèses de régularité de la fonction de lien). La variable à expliquer y est maintenant reliée à la variable explicative x par le biais d'une fonction de lien inconnue que l'on doit estimer :

$$y = f(x) + \varepsilon$$

Le théme commun de la régression fonctionnelle pour estimer $f(x)$ est l'idée d'un lissage local. Sa qualité dépend de la présence de suffisament de données dans le voisinage de x. Lorsque la variable explicative est unidimensionnelle, de nombreuses techniques de lissage sont disponibles : méthode des noyaux, splines de lissage, ... Mais lorsque la dimension de x devient importante, le nombre d'observations nécessaires pour le lissage local croît de manière exponentielle avec cette dimension. À moins de disposer d'un échantillon de taille gigantesque, ces méthodes d'estimation non paramétrique ne seront plus valides à cause de la faible abondance de données dans les régions d'intérêt (Big data). De plus, l'interprétation de la fonction de lien n'est pas toujours évidente.
Pour surmonter ce problème de dimension et d'interprétation de l'effet de la covariable x, des modéles de régression semi-paramétrique ont été developpés.

Modèles semi-paramétriques :

Les modèles semi-paramétriques ont alors été développés pour conjuguer les avantages des approches paramétriques et non paramétriques, à savoir la capacité d'interprétation des modéles paramétriques et la souplesse des modèles non paramètriques.
Dans de ce type de modèle, la variable à expliquer y dépend généralement de x par le biais d'un nombre fini de paramétres euclidiens $\theta'_1, \ldots, \theta'_k$ et d'un paramètre fonctionnel f.

Cet mémoire se compose de deux chapitres. Le premier chapitre est consacré sur les modèles semi-paramétriques.
Le dernier chapitre est sur estimation semi-paramétrique de la fonctions de distrubition dans le modèle à indice simple .

Chapitre 1

Sur les modèles semi-paramétriques

Nous commençons par citer le modèle partiellement linéaire qui est l'un des modèles semi-paramétriques les plus simples et les plus utilisés dans la pratique.

L'introduction de ce type de modèles semi-paramétriques nous assure une estimation simple car elle n'implique rien de plus que l'estimation de base du noyau des fonctions de régression avec la régression des moindres carré.
Ce modèle sert également à illustrer un certain nombre de questions un peux subtil qui se posent dans l'estimation des modèles semi-paramétriques en général.

Par exemple, le paramètre de dimension fini (la partie paramétrique du modèle) généralement peut être estimé avec le taux défaut de convergence paramétrique, bien que nous devons généralement des conditions de régularité importantes ainsi que des conditions plus strictes sur les paramètres de lissage afin d'obtenir le taux de convergence pour la partie paramétrique du modèle.

1.1 Les modéles partiellement linéaires :

Un modèle semi-paramétrique partiellement linéaire est donné par

$$Y_i = X_i'\beta + g(Z_i) + \varepsilon_i, \quad i = 1, \ldots, n \tag{1.1}$$

Où X_i est un $p \times 1$ vecteur, β est une $p \times 1$ vecteur des paramètres inconnus et $Z_i \in \mathbb{R}^q$. La forme fonctionnelle de $g(\cdot)$ n'est pas spécifiée. Le paramètre dimensionnel β constitue la partie paramétrique du modèle et la fonction inconnue $g(\cdot)$ représente la partie non paramétrique.

Les données sont supposées être i.i.d avec $\mathbb{E}(\varepsilon_i \backslash X_i, Z_i) = 0$, et nous permettons un processus d'erreur à condition d'irrégularité de la variance d'un variable aléatoire $\mathbb{E}(\varepsilon_i^2 \backslash x, z) = \sigma^2(x, z)$ de forme inconnue. Nous focalisons notre discussion sur la façon d'obtenir un estimateur de β $\sqrt{n}$-consistent, car une fois cela fait, un estimateur de $g(\cdot)$ peut être facilement obtenu.

1.1.1 Identification de β :

Certaines conditions d'identification seront nécessaires pour identifier le vecteur de paramètres β. Observer que X ne peut pas contenir une constante (i.e., β ne peut pas contenir une interception) parce que, s'il y a une interception, dites α, il n'est pas nécessaire d'identifier séparément la fonction inconnue $g(\cdot)$. Autrement dit, pour toute constante $c \neq 0$, nous observons

$$\alpha + g(z) = [\alpha + c] + [g(z) - c] \equiv \alpha_{new} + g_{new}(z)$$

donc la somme de la nouvelle ordonnée à l'origine et de la nouvelle fonction $g(\cdot)$ est équivalente à la somme de 1.1.

Puisque la forme fonctionnelle de $g(\cdot)$ n'est pas spécifiée, il s'agit immédiatement d'un terme d'interception qui ne peut être identifié dans le modèle partiellement linéaire.

Après l'obtention la distribution asymptotique de notre estimateur semi-paramétrique de β,

$$\Phi \stackrel{def}{=} \mathbb{E}\{[X - \mathbb{E}(X|Z)][X - \mathbb{E}(X|Z)]'\}$$

doit être une matrice définie positive, ce qui implique que X ne peut pas contenir une constante et qu'aucun des éléments de X ne peut être une fonction déterministe de Z , sinon $X - \mathbb{E}(X \backslash Z) = 0$ pour ce composant et Φ_i sera singulière.

1.1.2 L'estimateur de Robinson

Nous présentons d'abord un estimateur infaisable de 1.1 pour illustrer la mécanique impliquée dans l'estimation de β. Prenons l'expression de 1.1 conditionnelle à Z_i, nous obtenons

$$\mathbb{E}(Y_i|Z_i) = (X_i - \mathbb{E}(X_i|Z_i))'\beta + g(Z_i) \tag{1.2}$$

Après la soustraction de 1.2 par 1.1, elle nous donne

$$Y_i - \mathbb{E}(Y_i|Z_i) = (X_i - \mathbb{E}(X_i|Z_i))'\beta + \varepsilon_i \tag{1.3}$$

En définissant la notation abrégée $\tilde{Y}_i = Y_i - \mathbb{E}(Y_i|Z_i)$ et $\tilde{X} = X_i - \mathbb{E}(X_i \backslash Z_i)$, et en appliquant la méthode des moindres carrés à 1.3, on obtient un estimateur de β donné par :

$$\hat{\beta}_{inf} = \Big[\sum_{i=1}^{n} \tilde{X}_i \tilde{X}_i'\Big]^{-1} \sum_{i=1}^{n} \tilde{X}_i \tilde{Y}_i \tag{1.4}$$

Par le Théorème central limite de Linderberg-Levy [1] nous aurons immédiatement,

$$\sqrt{n}(\hat{\beta}_{inf} - \beta) \underset{d}{\longrightarrow} N(0, \Phi_i^{-1} \Psi_i \Phi_i^{-1}) \tag{1.5}$$

A condition que Φ_i soit défini positif, où $\Psi_i = \mathbb{E}[\sigma^2(X_i, Z_i)\tilde{X}_i\tilde{X}_i']$,
$\Phi_i = [\mathbb{E}(\tilde{X}_i\tilde{X}_i')]$.

L'idée fondamentale qui sous-tend cette procédure est d'abord éliminer la fonction inconnue $g(\cdot)$ ont en soustrayant 1.2 de 1.1. Bien que la fonction inconnue $g(\cdot)$ disparaisse dans 1.3, deux nouvelles fonctions inconnues sont introduites, à savoir $\mathbb{E}(Y_i|Z_i)$ et $\mathbb{E}(X_i|Z_i)$.

Donc, l'estimateur $\hat{\beta}_{inf}$ ci-dessus n'est pas possible parce que $\mathbb{E}(Y_i|Z_i)$ et $\mathbb{E}(X_i|Z_i)$ sont inconnus. Ces dernières peuvent être facilement estimées à l'aide des méthodes du noyau, donc nous pouvons remplacer les espérances conditionnelles inconnus qui apparaissent en version $\hat{\beta}_{inf}$ avec leur estimateurs à noyau , obtenant ainsi un estimateur possible de β.

Nous remplaçons $\tilde{Y}_i = Y_i - \mathbb{E}(Y_i|Z_i)$ et $\tilde{X}_i = X_i - \mathbb{E}(X_i|Z_i)$ par $Y_i - \hat{Y}_i$ et $X_i - \hat{X}_i$ respectivement, où

$$\hat{Y}_i \equiv \hat{\mathbb{E}}(Y_i|Z_i) \overset{def}{=} n^{-1} \sum_{j=1}^{n} Y_j K_h(Z_i, Z_j)/\hat{f}(Z_i) \tag{1.6}$$

$$\hat{X}_i \equiv \hat{\mathbb{E}}(X_i|Z_i) \overset{def}{=} n^{-1} \sum_{j=1}^{n} X_j K_h(Z_i, Z_j)/\hat{f}(Z_i) \tag{1.7}$$

et

$$\hat{f}(Z_i) = n^{-1} \sum_{j=1}^{n} K_h(Z_i, Z_j), \tag{1.8}$$

où

$$K_h(Z_i, Z_j) = \prod_{s=1}^{q} h_s^{-1} K\Big(\frac{Z_{is} - Z_{js}}{h_s}\Big)$$

La présence du dénominateur aléatoire $\hat{f}(Z_i)$ peut causer quelques difficultés techniques quand on décrit la distribution asymptotique de l'estimateur β. Pour cela, nous utilisons deux approches, la première utilise une fonction qui «réduit» les observations pour

1. LindebergLévy CLT. Soient X_1, X_2, une suite i.i.d. de variables aléatoires avec $E[X_i] = \mu$ et la $Var[X_i] = \sigma^2 < \infty$. Quand n s'approache de l'infini, les variables aléatoires $\sqrt{n}(S_n - \mu)$ converge en loi vers une variable aléatoire normale $N(0, \sigma^2)$ $\sqrt{n}((\frac{1}{n}\sum_{i=1}^{n} X_i) - \mu) \xrightarrow{d} N(0, \sigma^2)$

lesquelles le dénominateur est petit et l'autre utilise la méthode de la densité pondérée pour se débarrasser complètement du dénominateur aléatoire. On définit un estimateur réalisable de β par :

$$\hat{\beta} = \left\{\sum_i (X_i - \hat{X}_i)(X_i - \hat{X}_i)'\right\}^{-1} \sum_i (X_i - \hat{X}_i)(Y_i - \hat{Y}_i)\mathbb{1}_i \tag{1.9}$$

où

$$\mathbb{1}_i = \begin{cases} 1 & si \quad \hat{f}(Z_i) \geq b \\ 0 & \quad sinon \end{cases}$$

et où le paramètre de lissage $b = b_n > 0$ satisfait $b_n \longrightarrow 0$ si $n \longrightarrow \infty$.
Pour dériver la distribution asymptotique de $\hat{\beta}$ nous fournissons d'abord une définition et quelques hypothèses.
Nous utiliserons $\mathcal{G}_v^\alpha$, ou $\alpha > 0$; et $v \geq 2$ est un nombre entier. dénoter la classe de fonction lisse telle cela si $g \in \mathcal{G}_v^\alpha$ alors g est le temps v différentiable ; g et ses fonctions de dérivée partielle (à l'ordre v) tous satisfont quelques conditions de type Lipschitzien telles que $\mid g(z) - g(z)' \mid \leq H_g(z) \parallel z' - z \parallel$, où $H_g(z)$ est une fonction continue ayant le moment d'ordre α et où $\parallel \cdot \parallel$ dénotent la norme euclidienne,i.e.,

$$\parallel z \parallel = \sqrt{\sum_{j=1}^{q} z_j^2}.$$

Les hypothèses

- (i) $(Y_i, X_i, Z_i), i = 1, 2, \ldots, n$ sont des observations i.i.d, Z_i admet une fonction de densité de probabilité (PDF) $f \in \mathcal{G}_{v-1}^\infty$ (i.e. f est limité), $g(\cdot) \in \mathcal{G}_v^4$ où $v > 2$ est un nombre entier.

- (ii) $\mathbb{E}(u|X, Z) = 0, \mathbb{E}(u^2|x, z) = \sigma^2(x, z)$ est continu dans z, tant que X et u ont de quatrièmes moments finis.

- (iii) $K(\cdot)$ Est un noyau produit, le noyau univarié $k(\cdot)$ est un noyau d'ordre v limité, et $k(v) = O(1/[1+ \mid v \mid]^{v+1})$.

- (iv) Lorsque $n \longrightarrow \infty$,$n(h_1 \ldots h_q)^2 b^4 \longrightarrow \infty$, $nb^{-4} \sum_{s=1}^{q} h_s^{4v} \longrightarrow 0$

L'hypothése 2.1 (i) définit un ensemble de conditions de lissage et de moment.

La fonction inconnue $g(z)$ et $\mathbb{E}(X|z)$ sont supposées être différentiable d'ordre v. Celles-ci, avec un noyau d'ordre v du 2.1 (iii), assurent que le biais de l'estimateur à noyau est d'ordre $O(\sum_{s=1}^{q} h_s^v)$.
L'hypothése 2.1 (iv) est utilisée dans Robinson (1988)[42]. On peut ignorer le paramètre d'ajustement b dans l'application empirique, car on peut laisser $b \longrightarrow 0$ à un taux extrêmement lent. Ainsi, l'hypothése 2.1 (iv) est fondamentalement équivalente à

$$\sqrt{n}\Big[\sum_{s=1}^{q} h_s^{2v} + (nh_1 \dots h_q)^{-1}\Big] \quad \longrightarrow \quad 0 \quad quand \quad n \longrightarrow \infty$$

Cette condition est facile à comprendre, $O(\sum_{s=1}^{q} h_s^{2v} + (nh_1 \dots h_q)^{-1})$ est l'ordre du MSE non paramétrique. la différence entre l'estimateur $\hat{\beta}$ réalisable et l'estimateur infaisable $\hat{\beta}_{\text{inf}}$ est proportionnelle à la moyenne des erreurs d'estimateurs non paramétriques au carré. Par conséquent, pour que $\hat{\beta}$ soit un estimateur $\sqrt{n}$-consistent de β, il faut que les termes d'erreur d'estimation au carré soient plus petits que $n^{-\frac{1}{2}}$, ce qui donne l'hypothése 2.1 (iv).

Théorème 1.1. *sous les hypothése 2.1 ,en a*

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \Phi^{-1}\Psi\Phi^{-1}) \tag{1.10}$$

à condition que Φ *est définie positive, où*

$$\Phi = \mathbb{E}[\tilde{X}_i \tilde{X}_i'], \Psi = \mathbb{E}[\sigma^2(X_i, Z_i)\tilde{X}_i \tilde{X}_i'] \quad et \quad \tilde{X}_i = X_i - \mathbb{E}(X_i|Z_i)$$

la preuve du théorème(1.1) peut être trouvée dans Robinson (1988)[42].

1.1.3 Estimation de la composante non paramétrique :

D'aprés 1.2 nous connaissons cela

$$g(Z_i) = \mathbb{E}(Y_i - X_i'\beta|Z_i)$$

Donc après l'obtention $\hat{\beta}$, on donne à un estimateur de $g(z)$

$$\hat{g}(z) = \frac{\sum_{j=1}^{n}(Y_i - X_j'\hat{\beta})K_h(z, Z_j)}{\sum_{j=1}^{n} K_h(z, Z_j)} \tag{1.11}$$

Où $\hat{\beta}$remplacé par $\hat{\beta}_f$. nous savons que l'estimation de noyau non paramétrique a un taux de convergence qui est plus lent que le $\sqrt{n}-$taux parmétrique. Par conséquent, il est facile de voir que asymptotiquement, $\hat{g}(z)$ est équivalent à l'estimateur infaisable suivant qui utilise la vraie valeur de β :

$$\hat{g}(z) = \frac{\sum_{j=1}^{n}(Y_i - X_j'\beta)K_h(z, Z_j)}{\sum_{j=1}^{n}K_h(z, Z_j)} \tag{1.12}$$

le taux de convergence et la distribution asymptotique de $\hat{g}(z)$, à partir desquels on peut obtenir immédiatement la distribution asymptotique de $\hat{g}(z)$, sont discutés au $(Nonparametric\ \ econom$ $theory\ \ and\ \ practice\ \ /Qili\ \ \&jeffrey\ \ S.Racine,\ \ (57-106))$.
Notons que le choix de h_s pour estimer $g(z)$ peut être très différent de ceux pour estimer β afin d'obtenir $\hat{\beta}$, un noyau d'ordre supérieur est nécessaire si $q \geq 6$.

Cependant, lors de l'estimation de $g(z)$, il n'est pas nécessaire d'utiliser un noyau d'ordre supérieur quelle que soit la valeur de q.

Par conséquent, on peut toujours utiliser un noyau non négatif du second ordre pour estimer $g(z)$, et on pourrait choisir les paramètres de lissage par validation croisée des moindres carrés (en estimant $g(z)$), on peut toujours choisir $h(1), \ldots, h(q)$ pour minimiser

$$\sum_{i=1}^{n}[Y_i - X_i'\hat{\beta} - \hat{g}_{-i}(Z_i, h)]^2 \tag{1.13}$$

comme défini en 1.11 avec z remplacé par Z_i, et $\sum_{j=1}^{n}$ remplacé par $\sum_{j=1, j\neq i}^{n}$
Notons que dans 1.13 la variable dépendante est $Y_i - X_i'\hat{\beta}$ au lieu de Y_i. Car

$$\hat{\beta} - \beta = O_p(n^{-\frac{1}{2}})$$

a un taux de convergence plus rapide que n'importe quel estimateur non paramétrique, on peut remplacer $\hat{\beta}$ en 1.13 avec β pour étudier le comportement asymptotique de la validation croisée sélectionnée $\hat{h}_s's$. le taux de convergence de $\hat{h}_s's$ est le même que discutée au $(Nonparametric\ \ econometrics : theory\ \ and\ \ practice\ \ /Qili\ \ \&jeffrey\ \ S.Racine,\ \ (57-106))$.
On peut aussi sélectionner $\hat{\beta}$ et h simultanément en minimisant

$$\sum_{i=1}^{n}[Y_i - X_i'\beta - \hat{g}_{-i}(Z_i, h)]^2$$

Sous condition générale, y compris l'utilisation d'un noyau de second ordre, le choix de validation croisée de h sera de l'ordre $O_p(n^{-\frac{1}{(q+4)}})$. Cet ordre satisfait à les hypothéses 2.1 (iv), si $q \geq 3$.Donc, quand $q \leq 3$, on peut choisir simultanément les $h_s's$ et β en minimisant

$$\sum_{i=1}^{n}[Y_i - X_i'\beta - \hat{g}_{-i}(Z_i, h)]^2$$

Le $\hat{h}_s's$ résultant sera d'ordre $O_p(n^{-\frac{1}{(q+4)}})$, et le $\hat{\beta}$ résultant sera $\sqrt{n}-$consistent ayant la variance asymptotique donnée dans le théorème 1.1 On peut aussi utiliser des résultats d'expansion du second ordre (dans le $h_s's$) dans le modèle partiellement linéaire pour choisir les paramètres les plus faibles afin de minimiser l'estimation MSE(erreur quadratique moyenne) jusqu'au second ordre, voir Linton (1995).

1.2 Les modèles linéaires

Dans cette partie, nous considérons un autre modèle semi-paramétrique populaire, appelé l'indice simple. Ce modèle a été largement utilisé par les économétriciens appliqués.

Un modèle semi-paramétrique à indice simple est de la forme

$$Y = g(X'\beta_0) + \varepsilon \tag{1.14}$$

où Y est la variable dépendante, $X \in \mathbb{R}^q$ est le vecteur de la variable explicative, β_0 est le vecteur $q \times 1$ des paramètres inconnus, et ε est l'erreur qui satisfait $\mathbb{E}(\varepsilon|X) = 0$. Le terme $x'\beta_0$ est appelé un « indice simple » parce que c'est un scalaire (un seul indice) même si x est un vecteur.

La forme fonctionnelle de $g(\cdot)$ est inconnu. Ce modèle est semi-paramétrique dans la nature puisque la forme fonctionnelle d'indice linéaire est spécifiée, alors que $g(\cdot)$ n'est pas précisé .

Nous illustrons d'abord cet exemple populaire dans un modèle semi-paramétrique d'indice simple.

Dans le modèle de choix binaire, si l'on accepte l'indice linéaire paramétrique qui régit les choix mais ne veut pas spécifier la loi inconnue du terme d'erreur, alors on arrive à un modèle semi-paramétrique à indice simple.

Plus précisément, en considérant la relation entre une variable dépendante binaire (Y) et d'autres covariables (X), on pourrait modéliser cette relation comme suit :

$$Y_i = \begin{cases} 1 & si \quad Y^* \overset{def}{=} \alpha + X_i'\beta + \epsilon_i > 0 \\ 0 & si \quad Y^* \overset{def}{=} \alpha + X_i'\beta + \epsilon_i \leq 0 \end{cases} \tag{1.15}$$

où Y^* est une variable aléatoire. Notons ici, $\epsilon = Y^* - \mathbb{E}(Y^*|X)$, qui diffère de $\varepsilon = Y - \mathbb{E}(Y|X)$ défini en 1.14 parce que $Y \neq Y^*$.
Par exemple, Y pourrait représenter une décision de participation au marché du travail où si Y est égal à 1, un individu participe au marché du travail, alors que si Y est égal à 0, ce n'est pas.
Les variables explicatives X contiennent un ensemble de facteurs économiques qui pourraient influencer la décision de participer, tels que l'âge, l'état matrimonial, l'éducation, les antécédents professionnels et le nombre d'enfants.

Le modèle 1.15 suppose une fonction de lien paramétrique linéaire entre la décision de pratiquer ou non linéaire sur le marché du travail Y et les variables explicatives X. Nous sommes principalement intéressés par l'estimation de β, qui reflète l'impact des changements de X sur la probabilité de participer au marché du travail.

Les méthodes paramétriques pour l'estimation de β nécessitent de spécifier la loi (inconnue) du terme d'erreur ϵ. Une hypothèse populaire est que ϵ a une loi normale, c'est-à-dire $\epsilon \sim N(0, \sigma^2)$. On peut montrer que β et σ^2 ne peuvent pas être identifiés conjointement sans conditions d'identification supplémentaires (voir Maddala (1966)). Par exemple, si nous supposons que $\sigma = 1$, alors β est identifié, et nous pouvons utiliser l'estimation du maximum de vraisemblance pour estimer β.

Cependant, si le terme d'erreur ne possède pas une loi normale, alors l'approche paramétrique produira en général des estimations inconsistentes, c'est-à-dire, de $P(Y = 1|x) = \mathbb{E}(Y|x)$ soit $f_\epsilon(\cdot)$ le vrai CDF (Fonction de Distribution Cumulative) de ϵ. De 1.15 , en a

$$\begin{aligned}
\mathbb{E}(Y|x) &= \sum_{y=0.1} P(y|x) \\
&= 1 \times P(Y = 1|x) + 0 \times P(Y = 0|x) \\
&= P(Y = 1|x) \\
&= P(\alpha + x'\beta + \epsilon > 0) \\
&= P(\epsilon > -(\alpha + x'\beta)) \\
&= 1 - P(\epsilon < -(\alpha + x'\beta)) \\
&= 1 - F(-(\alpha + x'\beta)) \\
&\equiv m(\alpha + x'\beta)
\end{aligned}$$

où $f_\epsilon(\cdot)$ c'est le CDF de ϵ. Notons que si ϵ a une distribution symétrique, alors $F(\alpha + x'\beta) = 1 - F(-(\alpha + x'\beta))$, et nous avons $m(\cdot) = F(\cdot)$ Dans ce cas.

Par exemple, si $\epsilon \sim N(0,1)$ $(\sigma = 1)$, alors 1.15 devient de ce modèle :

$$\mathbb{E}(Y|x) = P(Y = 1|x) = \Phi(\alpha + x'\beta) \tag{1.16}$$

où $\Phi(\cdot)$ est le CDF d'une variable standard normale. Si, d'autre part, ϵ a la distribution logistique asymétrique, alors 1.15 résultats d'un modèle logistique de la forme

$$\mathbb{E}(Y|x) = P(Y = 1|x) = \frac{1}{1 + \exp(\alpha + x'\beta)} \tag{1.17}$$

À partir de 1.16 et 1.17, nous voyons que les différentes hypothèse distributionnelle pour ϵ conduire à différentes formes fonctionnelles pour la probabilité conditionnelle de $Y \ = 1$.

Par conséquent, une estimation paramétrique consistent de $P(Y = 1|x) = \mathbb{E}(Y|x)$ nécessite la spécification distributionnelle correcte de ϵ.

Un modèle semi-paramétrique d'indice simple évite donc le problème d'erreur de spécification de la distribution.

De plus, le modèle semi-paramétrique à indice simple 1.14 est plus général que le modèle à choix binaire puisque nous n'exigeons pas que la variable dépendante soit nécessairement de nature binaire.

Comme nous verrons que α dans la partie (2.3.1) de paramètre d'emplacement ne peut pas être identifié quand la forme fonctionnelle de $g(\cdot)$ est inconnu, c'est pourquoi nous écrivons notre modèle semi-paramétrique 1.14 seulement comme une fonction de $X_i'\beta_0$. Nous ceci discutons et d'autre condition d'identification dans la partie suivante .
Notons que le modèle 1.14 implique que $\mathbb{E}(Y|x) = g(x'\beta_0)$, d'où y dépend de x seulement par la combinaison linéaire $x'\beta_0$ et la relation est caractérisée par la fonction de lien $g(\cdot)$. Nous voudrions souligner ici que, comme le modèle partiellement linéaire, le modèle semi-paramétrique à indice simple est une approche alternative conçue pour atténuer des effets résultant de la malédiction de dimension.
Aussi, nous soulignons que Y peut être continu ou discret, c'est-à-dire, il n'y a aucune raison de limiter Y pour être une variable binaire.

1.2.1 Les conditions d'identification :

Pour le modèle semi-paramétrique à indice simple, on a

$$\mathbb{E}(Y|x) = g(x'\beta_0)$$

Ichimmura (1993)[28],Manski(1988)[33] et Horowitz (1988,pp.14-20)[26]
Fournir une excellente explication intuitive des conditions d'identifiabilité sous-jacentes aux modèles semi-paramétrique à indice simple (c'est-à-dire l'ensemble des conditions dans lesquelles le vecteur de paramètre inconnu β_0 et la fonction inconnue $g(\cdot)$ peuvent être raisonnablement estimés).

Nous discutons brièvement de ces conditions, puis les résumons dans la proposition.

D'abord,$g(\cdot)$ ne peut pas être une fonction constante, car il est évident que β_0 n'est pas identifié. Ensuite, comme dans le cas de la régression linéaire, les différentes composantes de x ne peuvent pas avoir une relation linéaire parfaite (multicollinéarité parfaite). Une autre restriction est que x contient au moins une variable aléatoire continue. Intuitivement, cela peut être compris par le biais du raisonnement suivant. Si x ne contient que des variables discrètes, disons quelques variables muettes 0-1, alors le support de x est fini, tout comme le support de la variable scalaire $v = x'\beta$ pour tous les vecteurs de β. Évidemment alors il existera un nombre infini de fonctions $g(\cdot)$ qui diffèrent en termes de leur β tel que $g(x'\beta) = \mathbb{E}(Y|x)$.
C'est tout simplement parce que $\mathbb{E}(Y|x) = g(x'\beta)$ n'impose qu'un nombre fini de restrictions à la fonction inconnue $g(\cdot)$, donc il existe un nombre infini de choix différents de $g(\cdot)$ et β qui satisfont à l'ensemble fini de restrictions imposées par $\mathbb{E}(Y|x) = g(x'\beta)$. Voir Horowitz (1988)[26] pour un exemple spécifique et une illustration de ce point.
Aussi, x ne peut pas contenir un constant. En d'autres termes, β_0 ne peut pas contenir de paramètre de localisation, et β_0 n'est identifiable que jusqu'à une échelle. Ce suit parce que, pour toutes les constantes non nulles α_1 et α_2 et pour toute fonction $g(\cdot)$ et β fixe, on peut toujours trouver une autre fonction, disons $g_2(\cdot)$, définie par $g_2(\alpha_1+\alpha_2 x'\beta) = g(x'\beta)$. Par conséquent, β n'est pas identifiée sans certaines restrictions de localisation et d'échelle (normalisation). Une normalisation courante est que x ne contient pas de constante, ce que l'on appelle la normalisation de l'emplacement. Pour la normalisation dite d'échelle, une approche consiste à supposer que le vecteur β a une longueur unitaire, c'est-à-dire $\|\beta\| = 1$, où

$$\|\beta\| = \{\sum_{j=1}^{q} \beta_i^2\}^{-\frac{1}{2}}$$

est la norme Euclidienne (longueur) de β. Une autre normalisation d'échelle populaire consiste à supposer que le premier composant de x a un coefficient unitaire, et cette première composante est supposée être une variable continue.
Nous summerize les conditions ci-dessus dans l'identification suivante :

1.2.1.1 Identification de modèle à indice simple

Pour le modèle semi-paramétrique à indice simple 1.14, l'identification de β_0 et $g(\cdot)$ nécessite que

- (i) x ne doivent pas contenir d'une constante (interception) , et x doit contenir au moins une variable continue. De plus $\|\beta_0\| = 1$
- (ii) $g(\cdot)$ est différentiable et n'est pas une fonction constante sur le support de $x'\beta_0$
- (iii) Pour les composants discrets de x, variant les valeurs du support de $x'\beta_0$ en sous-ensembles disjoints.

Nous avons déjà discuté comment, quand x ne contient que des variables discrètes, β_0 et $g(\cdot)$ ne sont pas identifiés.nous avons déjà discuté comment, quand x ne contient que des variables discrètes, β et $g(\cdot)$ ne sont pas identifiés. Cependant, quand $g(\cdot)$ est supposé être une fonction croissante, on peut obtenir des bornes identifiées sur les composants de β. Pour une discussion détaillée sur la façon de caractériser les limites lorsque toutes les composantes de x sont discrètes, voir Horowitz (1988, pp.17-20)[26]

1.2.2 Les Méthodes d'estimation :

1.2.2.1 La méthode d'Ichimmura :

Si la forme fonctionnelle de $g(\cdot)$ était connue, 1.14 deviendrait un modèle de régression non linéaire standard, et nous pourrions utiliser la méthode des moindres carrés non linéaires pour estimer β_0 en minimisant

$$\sum_{i=1}[Y_i - g(X_i'\beta)]^2 \tag{1.18}$$

par rapport à β dans le cas de $g(\cdot)$ est une fonction inconnue, il faut d'abord estimer $g(\cdot)$, nous devons d'abord estimer $g(\cdot)$. Cependant, la méthode du noyau n'estime pas $g(X_i'\beta_0)$ directement parce que non seulement $g(\cdot)$ est inconnu, mais aussi β_0, cependant, pour une valeur donnée de β on peut estimer

$$G(X'\beta) \overset{def}{=} \mathbb{E}(Y_i|X'\beta) = \mathbb{E}[g(X'\beta_0)|X'\beta] \tag{1.19}$$

par la méthode du noyau où la dernière égalité découle du fait que $\mathbb{E}(u_i|X_i'\beta) = 0$ pour tous β depuis $\mathbb{E}(u_i|X_i) = 0$.
Lorsque que $\beta = \beta_0$, $G(X'\beta_0) = g(X'\beta_0)$, alors qu'en général
$G(X_i'\beta) \neq g(X'\beta_0)$ si $\beta \neq \beta_0$. Un estimateur de noyau non-paramétrique (leave-one-out) de $G(X_i'\beta)$ est donné par

$$\hat{G}_{-i}(X_i'\beta) \equiv \hat{\mathbb{E}}_{-i}(Y_i|X_i'\beta) = \frac{(nh)^{-1}\sum_{j=1,j\neq i}^{n} Y_j K\left(\frac{X_j'\beta - X_i'\beta}{h}\right)}{\hat{p}_{-i}(X_i'\beta)} \tag{1.20}$$

où

$$\hat{p}_{-i}(X_i'\beta) \quad = \quad (nh)^{-1}\sum_{j=1,j\neq i}^{n} K\left(\frac{X_j'\beta - X_i'\beta}{h}\right)$$

Ichimura (1993) suggère d'estimer $g(X_i'\beta_0)$ dans 1.15 par $\hat{G}_{-i}(X_i'\beta)$ de 1.17 et de choisir β par les moindres carrés semi-paramétriques non linéaires .
Il y a cependant un problème technique, à savoir que 1.20 a un dénominateur aléatoire, c'est-à-dire

$$\hat{p}(X_i'\beta) = (nh)^{-1}\sum_{j=1,j\neq i}^{n} K((X_j'\beta - X_i'\beta)/h).$$

Ichimura utilise une fonction de réduction pour couper les petites valeurs de $\hat{p}(X_i'\beta)$. Soit $p(x'\beta)$ le PDF(fonction de densité de probabilité) de $X_i'\beta$. Definie A_δ et A_n comme étant les suivants :

$$A_\delta = \{x : p(x'\beta) \geq \delta, \ \ \forall \ \ \beta \in \mathcal{B}\}$$

où $\delta > 0$ est une constante, $\mathcal{B}$ est un sous-ensemble compact dans $\mathbb{R}$, et

$$A_n = \{x : \|x - x^*\| \leq 2h, \forall x^* \in A_\delta\}$$

L'ensemble A_δ assure que le dénominateur de 1.17 ne soit pas trop proche de zéro pour $x \in A_\delta$. l'ensembleA_n est légèrement plus grand que A_δ, mais comme $n \longrightarrow \infty$, $h \longrightarrow 0$ et A_n rétrécit à A_δ
Ichimura (1993)[27] suggère de choisir β en minimisant la fonction d'objectif suivant :

$$S_n(\beta) = \sum_{i=1}^{n}\left[Y_i - \hat{G}_{-i}(X'\beta)\right]^2 \omega(X_i)\mathbb{1}(X_i \in A_n) \tag{1.21}$$

où $\hat{G}_i(X_i'\beta)$ est défini dans 1.20, $\omega(X_i)$ est une fonction de poids non négatif, et $\mathbb{1}(\cdot)$ est une fonction de l'indicatrice. Ceci est la fonction de réduction qui est :

$$\mathbb{1}(\cdot) = \begin{cases} 1 & si \quad X_i \in A_n \\ 0 & \quad sinon \end{cases}$$

La fonction de réduction assure que le dénominateur aléatoire dans l'estimateur noyau est positif avec la haute probabilité afin de simplifier l'analyse asymptotique.
Soit $\hat{\beta}$ l'estimation semi-paramétrique de β_0 obtenue par minimisation de 1.21 pour dériver l'asymptotique de $\hat{\beta}$, les conditions suivantes sont :

- $A(1)$: l'ensemble A_δ est compact, et la fonction de poids $w(\cdot)$ est bornée et positive sur A_δ. Définir l'ensemble $D_z = \{z : z = x'\beta, \beta \in A_\delta\}$. Soit $p(\cdot)$ indiquer le PDF (fonction de densité de probabilité) de $z \in D_z$, $p(\cdot)$ est délimité ci-dessous par une constante positive pour tout $z \in D_z$.
- $A(2)$: $g(\cdot)$ et $p(\cdot)$ sont trois fois différentiables par rapport à $z = x'\beta$. les troisièmes dérivées sont Lipschitz continues uniformément sur $\mathfrak{B}$ pour tout $z \in D_z$.
- $A(3)$: La fonction du noyau est un noyau borné de second ordre ayant un support borné, est deux fois différentiable, et sa dérivée seconde est Lipschitz continue.
- $A(4)$: $\mathbb{E}|Y^m| < \infty$ pour tout $m \geq 3$, $cov(Y|x)$ est borné et borné à partir de zéro pour tout $x \in A_\delta$

$$q \ln(h)/[nh^{3+3/(m-1)}] \longrightarrow 0$$

Ichimura (1993)[27] a prouvé les résultats suivants :

Théorème 1.2. *Sous les hypothèses A(1) à A(4)*

$$\sqrt{n}(\hat{\beta} - \beta_0) \longrightarrow N(0, \Omega_I) \quad en \quad loi$$

où $\Omega_I = V^{-1}\Sigma V^{-1}$, *et*

$$\begin{aligned} \Sigma &= \mathbb{E}\Big\{\omega(X_i)\sigma^2(X_i)\big(g_i^{(1)}\big)^2(X_i - \mathbb{E}_A(X_i - \mathbb{E}_A(X_i|X_i'\beta))) \\ &\times \; (X_i - \mathbb{E}_A(X_i|X_i'))'\Big\} \end{aligned}$$

Avec $g_i^{(1)} = [\delta g(v)/\delta v]|_{v=X_i'\beta_0}$, $\mathbb{E}_A(X_i|v) = \mathbb{E}_A(X_i|x_A'\beta_0 = v)$
avec x_A *la distribution de* X_i *à condition de* $X_i \in A_\delta$, *et*

$$V = \mathbb{E}\Big[\omega(X_i)\Big(g_i^1\Big)^2(X_i - \mathbb{E}_A(X_i|X_i'\beta_0))(X_i - \mathbb{E}_A(X_i|X_i'\beta_0))'\Big]$$

un estimateur consistent Ω_I *est donné par*

$$\hat{\Omega}_I = \hat{V}^{-1}\hat{\Sigma}\hat{V}^{-1}$$

où

$$\hat{V} = n^{-1}\sum_i \omega(X_i)\hat{g}^{(1)}(X_i'\hat{\beta})(X_i - \hat{\mathbb{E}}(X_i|X_i'\beta))(X_i - \hat{\mathbb{E}}(X_i|X_i'\beta))'$$

$$\hat{\Sigma} = n^{-1}\sum_i \omega(X_i)\hat{u}_i^2\hat{g}^{(1)}(X_i')(X_i - \hat{\mathbb{E}}(X_i|X_i'\beta))(X_i - \hat{\mathbb{E}}(X_i|X_i'\beta))'$$

$$\hat{u}_i = Y_i - \hat{g}(X_i'\hat{\beta}), \hat{g}^{(1)}(X_i'\hat{\beta}) = [\delta\hat{g}_{-i}(X_i'\beta)/\delta\beta]|_{\beta=\hat{\beta}}, \hat{g}_{-i}(X_i'\beta)$$

est défini dans 1.20

$$\hat{\mathbb{E}}(X_i|X_i'\beta)' = \sum_j X_j K((X_i - X_j)'\hat{\beta}) / \sum_j K((X_i - X_j)'\hat{\beta})$$

La preuve de Théorème 1.2 trés technique, pour plus de detail voir Ichimura (1993)[27].

1.2.2.2 Estimateur de Klein et Spady (1993) :

Lorsque le modèle d'indice simple est dérivé du modèle de choix binaire 1.15, et sous l'hypothèse que ϵ_i et X_i sont indépendants , Klein et Spady a suggéré l'estimation β par des méthodes de maximum de vraisemblance. L'estimation de la fonction de log −vraisemblance est

$$\mathcal{L}(\beta) = \sum_i (1 - Y_i) \ln(1 - \hat{g}(X_i'\beta)) + \sum_i Y_i \ln[\hat{g}(X_i'\beta)] \tag{1.22}$$

où $\hat{g}(X_i')$ est défini en 1.20.
La maximisation de 1.22 par rapport à β conduit à l'estimateur semi-paramétrique du maximum de vraisemblance de β, disons β_{KS}, proposé par Klien et Spady. Comme l'estimateur d'Ichimura (1993), la maximisation doit être effectuée numériquement en résolvant la condition de premier ordre obtenue à partir de 1.22.
Dans quelques conditions de régularité, y compris la présentation d'une fonction de réduction pour couper d'observations près de la frontière du support de X_i et l'utilisation de noyaux d'ordre plus hauts, Klien et Spady (1993) a montré que $\hat{\beta}_{KS}$ est $\sqrt{n}$−consistent et a une distribution asymptotique normale donnée par

$$\sqrt{n}(\hat{\beta}_{KS} - \beta) \longrightarrow N(0, \Omega_{KS})$$

où

$$\Omega_{KS} = \left[\mathbb{E}\left\{\frac{\delta p}{\delta\beta}\left(\frac{\delta p}{\delta\beta}\right)'\left[\frac{1}{p(1-p)}\right]\right\}\right]^{-1}$$

et

$$p = p(\epsilon < x'\beta) = F_{\epsilon|x}(x'\beta)$$

où $F_{\epsilon|x}(\cdot)$ est le CDF de ϵ_i conditionnel à $X_i = x$.
Klien et Spady (1993) ont également montré que l'estimateur proposé est semi-paramétriquement efficace en ce sens que la variance asymptotique de leur estimateur atteint la limite d'efficacité semiparamétrique.

Nous comparons la variance asymptotique de l'estimateur semi-paramétrique, Ω_{KS}, avec la contrepartie paramétrique Ω_{nls}. Le modèle paramétrique a deux paramètres supplémentaires, $\eta = (\gamma_0, \gamma_1)'$, Partitionnant le paramètre η en $\eta = (\gamma_0, \gamma_1)'$, puis la variance asymptotique pour $n^{\frac{1}{2}(\hat{\beta}_{nls}-\beta)}$ est

$$V_{nls} \stackrel{def}{=} (\mathcal{I}^0_{\beta\beta} - \mathcal{I}^0_{\beta\eta}(\mathcal{I}^0)^{-1}\mathcal{I}^0_{\eta\beta})^{-1}$$

En comparant cela avec V_{KS}^{-1}, on peut montrer que (e.g.,Pagan et Ullah (1999, p.278)) si

$$\mathbb{E}(X_i|X_i'\beta) = c_0 + c_1(X_i'\beta)$$

où c_0 et c_1 sont deux $q \times 1$ vecteur de constantes, alors

$$V_{KS}^{-1} = V_{nls}^{-1} \ \ ou \ \ quivalent \ \ , V_{KS} = V_{nls}$$

C'est-à-dire que l'estimateur semi-paramétrique est aussi efficace que l'estimateur paramétrique non linéaire des moindres carrés basé sur la forme fonctionnelle vraie connue $g(\cdot)$ lorsque $\mathbb{E}(X_i|X_i'\beta)$ est linéaire dans $X_i'\beta$ ("efficacité du premier ordre"). C'est semblable au cas modèle partiellement linéaire.
Cependant, lorsque $\mathbb{E}(X_i|X_i'\beta)$ n'est pas une fonction linéaire de $X_i'\beta$, on peut montrer que $V_{ks} - V_{mle}$ est définie positive. Par conséquent, l'estimateur semiparamétrique est asymptotiquement moins efficace que l'estimateur paramétrique non linéaire des moindres carrés basé sur la vraie fonction $g(\cdot)$.
La variance asymptotique $V_{ks} - V_{mls}$ est définie positive à moins que

$$\mathbb{E}(X_i|X_i'\beta) = c_0 + c_1(X_i'\beta)$$

De plus, dans ce cas, le résultat ne peut pas être amélioré puisque V_{ks} atteint déjà la limite d'efficacité semiparamétrique. La perte d'efficacité du modèle semi-paramétrique comparé à un estimateur paramétrique non linéaire des moindres carrés survient parce que la forme fonctionnelle de $g(\cdot)$ (ou, équivalent, $F_{\epsilon|x}(\cdot)$) est inconnue. Bien sûr, en partie, la forme fonctionnelle vraie $g(\cdot)$ typiquement inconnue, et dans ce cas, l'estimateur semi-paramétrique est robuste à la spécification de forme fonctionnelle de $g(\cdot)$.

Chapitre 2

Estimation semi-paramétrique de la fonctions de distrubition dans le modèle à indice simple

2.1 Introduction :

Les modèles d'indices fonctionnels simples ont reçu une attention considérable à cause de leurs larges applications dans beaucoup de domaines comme l'économie, la médecine, financier économétrique et ainsi de suite.

L'étude de ces modèles a été développée rapidement, voir l'Ait-Saidi et al. (2005[1], 2008a[3], 2008b[4]). Récemment, Attaoui et al. (2011)[42] ont étudié l'estimateur du noyau de la densité conditionnelle d'une variable de réponse scalaire Y, étant donné une variable aléatoire d'hilbertienne X quand les observations sont d'un modèle d'indice fonctionnel simple.
La convergence point à point et la convergence presque complète de l'estimateur avec les taux de ce modèle ont été obtenues pour des observations indépendantes.
En outre, $Ling \quad et \quad al.$(2012)[39] obtenu la normalité asymptotique de l'estimateur de la densité conditionnelle et le mode conditionnel pour l'estimateur $\alpha - mlange$ fonctionnel de la dépendance des données de séries chronologiques.
$Ling \quad et \quad al.$(2014)[40] a examiné la consistance presque complète la convergence point à point et la convergence presque complète uniforme de l'évaluation de l'estimation de noyau avec taux pour la densité conditionnelle dans l'arrangement(la mise) du $\alpha - mlange$ de fonctionnel des données, qui prolongent le cas i.i.d dans Attaoui et al.(2011)[42] à la dépendance

Arrangement, le taux de convergence de l'estimation du noyau pour le mode conditionnel a été aussi obtenu.
La principale contribution de ce chapitre est d'établir la normalité asymptotique de l'estimateur de la fonction de distribution conditionnelle dans le cas i.i.d.
lorsque le modèle à indice simple fonctionnel est fixé. En tant qu'application, l'asymptotique d'intervalle de confiance $(1-\Upsilon)$ pour la fonction de densité conditionnelle $F(\theta, y, x)$ est présenté.
Dans ce chapitre ,nous présentons le modèle, de même que des hypothèses de base qui sont nécessaires pour calculer le résultat principal. Nous énonçons le résultat principal de l'étude ; la normalité asymptotique de l'estimateur de la fonction de distribution conditionnelle.
En tant qu'application, l'asymptotique d'intervalle de confiance $(1-\Upsilon)$ de la fonction de distribution conditionnelle est donnée pour $0 < \Upsilon < 1$. Enfin, les preuves sont liées à dernière partie de ce chapitre.

2.2 Le modèle et quelques hypothèses de base

Soit $\{(X_i, Y_i), 1 \leq i \leq \eta\}$ être n variables aléatoires, identiquement distribué comme le couple aléatoire (X, Y) avec des valeurs dans $\mathcal{H} \times \mathbb{R}$, où $\mathcal{H}$ est un espace de Hilbert réel séparable avec la norme $\|\cdot\|$ générée par un produit interne $< \cdot, \cdot >$. Sous une telle structure topologique et pour θ une fonction fixe, on suppose que la distribution de probabilité conditionnelle de Y est donnée $< X, \theta > = < x, \theta >$ existe et est donné par

$$\forall y \in \mathbb{R}, F(\theta, y, x) = \mathbb{P}(Y \leq y| < X, \theta > = < x, \theta >). \tag{2.1}$$

L'estimateur à noyau non paramétrique $\hat{F}(\theta, y, x)$ de $F(\theta, y, x)$ est défini comme suit :

$$\hat{F}(\theta, y, x) = \frac{\sum_{i=1}^{n} K(h_k^{-1}(< x - X_i, \theta >)) H(h_H^{-1}(y - Y_i))}{\sum_{i=1}^{n} K(h_k^{-1}(< x - X_i, \theta >))} \tag{2.2}$$

Où K est un noyau , H est une fonction de distribution cumulative (cdf) et $h_k = h_{k,n}(resp, h_H = h_{H,n})$ est un ordre des nombres réels positifs qui va au 0 tant que $n \longrightarrow \infty$, et avec la convention $O/O = O$.
Pour tout $x \in \mathcal{H}, i = 1, \ldots, n$ et $y \in \mathbb{R}$

$$K_i(\theta, x) := K(h_k^{-1}| < x - X_i, \theta > |), \quad et \quad H_i(y) := H(h_H^{-1}(y - Y_i)).$$

Nous dénotons par $B_\theta(x, h) = \{X \in \mathcal{H} | O < | < x - X, \theta > | < h\}$ la boule centrée en x avec rayon h, soit N_x est un voisinage fixe de x en $\mathcal{H}$, S_R sera un sous-ensemble compact

fixe de $\mathbb{R}$.
Maintenant, nous présentons les suppositions suivantes de base :

(H1) $\mathbb{P}(X \in B_\theta(x, h_k)) =: \Phi_{\theta,x}(h) > 0, \quad \Phi_{\theta,x}(h) \longrightarrow 0$ si $h \longrightarrow 0$

(H2) La fonction de distribution cumulative $F(\theta, y, x)$ satisfait la condition de Hölder , c'est :
$\forall(y_1, y_2) \in S_R \times S_R, \forall(x_1, x_2) \in N_x \times N_x$.

$$|F(\theta, y_1, x_1) - F(\theta, y_2, x_2)| \leq C_{\theta,x}(\| x_1 - x_2 \|^{b_1} + |y_1 - y_2|^{b_2}), b_1 > 0, b_2 > 0.$$

(H3) Pour $j = 0, 1, H^j$ satisfait les conditions de lipschitz et

$$m =: \inf_{t\in[0,1]} K(t)H'(t) > 0$$

avec

$$\int H'(t)dt = 1 \ , \ \int H^2(t)dt < \infty \quad \text{et} \quad \int |t|^{b_2} H^1(t)dt < \infty$$

(H4) Pour Le noyau K est positive, avec le support compact $[0, 1]$ de classe $\mathcal{C}^1$ sur $[0, 1)$ tel que $K(1) > 0$ et sa dérivée K' existe sur $[0, 1)$ et $K' < 0$.
(H5) Pour chaque $u \in [0, 1]$,

$$\lim_{h\longrightarrow 0} \frac{\Phi_{\theta,x}(uh)}{\Phi_{\theta,x}(h)} = \lim_{h\longrightarrow 0} \xi_h^{\theta,x}(u) = \xi_h^{\theta,x}(u)$$

(H6) La bande passante h_H satisfait,

(i) $\frac{\log n}{n\Phi_{\theta,x}(h_K)} \longrightarrow 0$, si $n \longrightarrow \infty$.

(ii) $nh_H^2 \Phi_{\theta,x}^2(h_K) \longrightarrow \infty$, et $\frac{nh_H^3 \Phi_{\theta,x}(h_K)}{\log^2 n} \longrightarrow \infty$ si $n \longrightarrow \infty$.

(iii) $nh_H^2 \Phi_{\theta,x}^3(h_K) \longrightarrow 0$ si $n \longrightarrow \infty$.

(H7)

(i) $\frac{\Phi_{\theta,x}(h)}{n} + \Phi_x(h) = \mathcal{O}(\frac{1}{n})$

(ii) $\sqrt{n\Phi_{\theta,x}(h)} \longrightarrow 0$ si $n \longrightarrow \infty$.

Remarque 2.2.1. *L'hypothèse (H5) est connue comme (pour petit h) "l'hypothèse de concentration agissant sur la distribution de X" dans les espaces de dimension infinie.*
La fonction $\xi_h^x(\cdot)$ intervenant dans l'hypothèse (H7) est croissante pour tout h fixé. Sa limite ponctuelle $\xi_h^x(\cdot)$ joue un rôle déterminant.
Il est possible de spécifier cette fonction (avec $\xi_0(u) := \xi_0^x(u)$ dans les exemples ci-dessus par :

1. $\xi_0(u) = u^\gamma$
2. $\xi_0(u) = \delta_1(u)$, *où* $\delta_1(\cdot)$ *est une fonction de Dirac.*
3. $\xi_0(u) = \mathbb{1}_{]0,1]}(u)$.

2.2.1 Les résultats Principaux

2.2.1.1 Normalité asymptotique de l'estimateur $\hat{F}(\theta, y, x)$

Dans cette partie, nous donnons la normalité asymptotique de la fonction de distribution cumulative conditionnelle dans le modèle de l'indice simple fonctionnel .
Le principal résultat est donné dans le théorème suivant.

Théorème 2.1. *Sous Les hypothèses (H1)-(H7) nous avons*

$$\sqrt{\frac{n\Phi_{\theta,x}(h_K)}{\sigma^2(\theta, y, x)}}(\hat{F}(\theta, y, x) - F(\theta, y, x)) \xrightarrow{D} N(0, 1).$$

où

$$\sigma^2(\theta, y, x) = \frac{C_2(\theta, x)F(\theta, y, x)(1 - F(\theta, y, x))}{C_1^2(\theta, x)}$$

avec
$C_{\theta,x} = K^j(1) - \int_0^1 sK'(s)\beta_{\theta,x}(s)ds$ *pour* $j = 1, 2,$ "$\xrightarrow{D}$" *désigne la convergence dans la distribution.*

2.2.1.2 Preuve :

considérer, pour $i = 1, \ldots, n$

$$K_i(\theta, x) = K(h_K^{-1}(< x - X_i, \theta >)), H_i(y) = H\big(h_H^{-1}(y - Y_i)\big),$$

$$\hat{F}_N(\theta, y, x) = \frac{1}{n\mathbb{E}(K_1(\theta, x))}\sum_{i=1}^{n} K_i(\theta, x)H_i(y),$$

$$\hat{F}_D(\theta,x)=\frac{1}{n\mathbb{E}(K_1(\theta,x))}\sum_{i=1}^{n}K_i(\theta,x),$$

$$\Delta_i(x,\theta)=\frac{K(h_K^{-1}(< x-X_i,\theta >))}{\mathbb{E}K_1(\theta,x)}$$

Afin d'établir la normalité asymptotique de $\hat{F}(\theta,y,x)$, nous devons considérer la décomposition suivante :

$$\begin{aligned}\hat{F}(\theta,y,x)-F(\theta,y,x) &= \frac{\hat{F}_N(\theta,y,x)}{\hat{F}_D(\theta,x)}-\frac{C_1(\theta,x)F(\theta,y,x)}{C_1(\theta,x)}\\
&= \frac{1}{\hat{F}_D(\theta,x)}(\hat{F}_N(\theta,y,x)-\mathbb{E}\hat{F}_N(\theta,y,x))\\
&- \frac{1}{\hat{F}_D(\theta,x)}\big(C_1(\theta,x)F(\theta,y,x)-\mathbb{E}\hat{F}_N(\theta,y,x)\big)\\
&+ \frac{F(\theta,y,x)}{\hat{F}_D(\theta,x)}\big(C_1(\theta,x)-\mathbb{E}\big[\hat{F}_D(\theta,x)\big]\big)\\
&- \frac{F(\theta,y,x)}{\hat{F}_D(\theta,x)}\big(\hat{F}_D(\theta,x)-\mathbb{E}\hat{F}_D(\theta,x)\big)\end{aligned}$$

$$= \frac{1}{\hat{F}_D(\theta,x)}A_n(\theta,y,x)+B_n(\theta,y,x) \tag{2.3}$$

où

$$A_n(\theta,y,x)=\frac{1}{n\mathbb{E}K_1(x,\theta)}\sum_{i=1}^{n}\big\{(H_i(y)-F(\theta,y,x))K_i(\theta,x)$$

$$-\mathbb{E}\big[(H_i(y)-F(\theta,y,x))K_i(\theta,x)\big]\big\}=\frac{1}{n\mathbb{E}K_1(x,\theta)}\sum_{i=1}^{n}N_i(\theta,y,x)$$

et

$$N_i(\theta,y,x)=(H_i(y)-F(\theta,y,x))K_i(\theta,x)-\mathbb{E}[(H_i(y)-F(\theta,y,x))K_i(\theta,x)].$$

Il s'ensuit que,

$$\begin{aligned}n\Phi_{\theta,x}(h_k)Var(A_n(\theta,t,x)) &= \frac{\Phi_{\theta,x}(h_k)}{\mathbb{E}^2K_1(x,\theta)}Var(N_1)\\
&+ \frac{\Phi_{\theta,x}(h_k)}{\mathbb{E}^2K_1(x,\theta)}\sum_{|i-j|}\sum^{n}Cov(N_i,N_j)\end{aligned}$$

$$= V_n(\theta,t,x) \tag{2.4}$$

Alors, le reste de la preuve est basé sur les Lemmes suivants

Lemme 2.1. *Sous les hypothèses (H1)-(H3),(H5)et(H7),si $n \longrightarrow \infty$ nous avons*

$$n\Phi_{\theta,x}(h_k)Var(A_n(\theta,t,x)) \longrightarrow V(\theta,t,x)$$

où

$$V(\theta,t,x) = \frac{C_1(\theta,x)}{(C_2(\theta,x))^2}F(\theta,y,x)(1-F(\theta,y,x))$$

Lemme 2.2. *Sous les hypothèses (H1)-(H3),(H5)et(H7),si $n \longrightarrow \infty$ nous avons*

$$\sqrt{\left(\frac{n\Phi_{\theta,x}(h_k)}{V(\theta,y,x)}\right)} \qquad A_n(\theta,y,x) \xrightarrow{D} N(0,1)$$

où $\xrightarrow{D}$ dénote la convergence dans la distribution.

Lemme 2.3. *Conformément à suppositions (H1)-(H3),(H5)et(H7), si $n \longrightarrow \infty$ nous avons*

$$\sqrt{n\Phi_{\theta,x}(h_k)}B_n(\theta,y,x) \longrightarrow 0 \quad \textit{en probabilité.}$$

Maintenant, parce que les fonctions inconnues $C_j(\theta,x)$ et $F(\theta,y,x)$ intervenant dans l'expression de la variance, nous devons estimer les quantités $C_1(\theta,x)$,$C_2(\theta,x)$ et $F(\theta,y,x)$, respectivement.

Par les hypothèses (H1)-(H4) nous savons que $a_j(\theta,x)$ peut être estimée par $\widehat{C}_j(\theta,x)$ qui est défini comme

$$\widehat{C}_j(\theta,x) = \frac{1}{n\Phi_{\theta,x}(h_k)}\sum_{i=1}^{n}K_i^j(\theta,x), j=1,2$$

$$\widehat{\Phi}_{\theta,x} = \frac{1}{n}\sum_{i=1}^{n}I_{\{|<x-X_i,\theta|>|<h_K\}}.$$

En appliquant l'estimateur à noyau de $F(\theta,y,x)$ donné ci-dessus, la quantité $\sigma^2(\theta,x)$ peut être estimé finalement par :

$$\hat{\sigma}^2(\theta,x) = \frac{\widehat{C}_2(\theta,x)\widehat{F}(\theta,y,x)}{\widehat{C}_1^2(\theta,x)}\int H^2(t)dt.$$

Ensuite, nous pouvons déduire le corollaire suivant :

Corollaire 2.1. *Sous les hypothèses du théorème (2.1), nous avons*

$$\sqrt{\frac{n\widehat{\Phi}_{\theta,x}(h_K)}{\widehat{\sigma}^2(\theta,y,x)}}(\widehat{F}(\theta,y,x)-F(\theta,y,x)) \xrightarrow{D} N(0,1).$$

Ainsi, après ce corollaire nous pouvons rapprochant $(1-\gamma)$ intervalle de confiance de $F(\theta,y,x)$ par

$$\widehat{F}(\theta,y,x) \pm t_{\gamma/2} \times \frac{\widehat{\sigma}(\theta,x)}{n\widehat{\Phi}_{\theta,x}(h_K)}$$

où est $t_{\gamma/2}$ le supérieur $\gamma/2$ quantile de normale standard N(0,1)

2.2.2 Preuves de lemmes techniques :

2.2.2.1 Preuve de lemme 2.1 :

Soit

$$V_n(\theta,y,x) = \frac{\Phi_{\theta,x}(h_k)}{\mathbb{E}^2K_1(\theta,x)}\mathbb{E}\Big[K_1^2(\theta,x)(H_1(y)-F(\theta,y,x))^2\Big]$$

$$= \frac{\Phi_{\theta,x}(h_k)}{\mathbb{E}^2K_1(\theta,x)}\mathbb{E}\Big[K_1^2(\theta,x)\mathbb{E}\Big((H_1(y)-F(\theta,y,x))^2| < \theta, X_1 > \Big)\Big] \quad (2.5)$$

En utilisant la définition de la variance conditionnelle, nous avons

$$\mathbb{E}\Big[\Big(H(h_H^{-1}(y-Y_1))-F(\theta,y,x)\Big)^2| < \theta, X_1 > |\Big] = J_{1n}+J_{2n}$$

où

$$J_{1n} = Var\Big(H(h_H^{-1}(y-Y_1))| < \theta, X_1 > \Big)$$

et

$$J_{2n} = \Big[\mathbb{E}\Big(H(h_H^{-1}(y-Y_1))| < \theta, X_1 > \Big) - F(\theta,y,x)\Big]^2$$

✓ **Concernant** J_{1n}

$$\begin{aligned} J_{1n} &= \mathbb{E}\Big[H^2\Big(\frac{y-Y_1}{h_H}| < \theta, x > \Big)\Big] - \Big(\mathbb{E}\Big[H^2\Big(\frac{y-Y_1}{h_H}\Big)| < \theta, x > \Big]\Big)^2 \\ &= \mathcal{J}_1 + \mathcal{J}_2 \end{aligned}$$

• par la propriété de double espérance conditionnelle, nous obtenons que

$$\begin{aligned} \mathcal{J}_1 &= \mathbb{E}\Big[H^2\Big(\frac{y-Y_1}{h_H}\Big)| < \theta, X_1 > \Big] \\ &= \int_{\mathbb{R}} H^2\Big(\frac{y-v}{h_H}\Big)dF(\theta,v,X_1) \\ &= \int_{\mathbb{R}} H^2(t)dF(\theta,y-h_Ht,X_1) \end{aligned} \quad (2.6)$$

D'autre part, en intégrant par partie et sous hypothèse (H3), nous avons avoir

$$\begin{aligned} \mathcal{J}_1 &= \int_{\mathbb{R}} 2H(t)H'(t)F(\theta,y-h_Ht,X_1)du \\ &= \int_{\mathbb{R}} 2H(t)H'(t)F(\theta,y-h_Ht,X_1-F(\theta,y,x))du \\ &+ \int_{\mathbb{R}} 2H(t)H'(t)F(\theta,y,x)du \end{aligned}$$

De toute évidence, nous avons

$$\int_{\mathbb{R}} 2H(t)H'(t)F(\theta,y,x)du = [H^2(t)F(\theta,y,x)]_{-\infty}^{+\infty} = F(\theta,y,x) \tag{2.7}$$

Ainsi

$$\int_{\mathbb{R}} H^2(t)dF(\theta,y-h_Ht,X_1) = F(\theta,y,x) + \mathcal{O}(h_k^{b_1} + h_H^{b_2}) \tag{2.8}$$

✓ **Concernant** $\mathcal{J}_{2n}$

$$\begin{aligned}
I &= \mathbb{E}(H_i(y)| < X_1, \theta >) \\
\mathbb{E}\Big(H\Big(\frac{y-Y_1}{h_H}\Big)| < X_1, \theta > \Big) &= \int_{\mathbb{R}} H\Big(\frac{y-u}{h_H}\Big) f(\theta,y,X_1)du, \\
&= \int_{\mathbb{R}} H\Big(\frac{y-u}{h_H}\Big) dF(\theta,y,X_1), \\
&= \int_{\mathbb{R}} H'\Big(\frac{y-u}{h_H}\Big) F(\theta,y,X_1)du, \\
&= \int_{\mathbb{R}} H'(t)(F(\theta,y-h_Ht,X_1) - F(\theta,y,x))dt \\
&+ F(\theta,y,x)) \int_{\mathbb{R}} H'(t)dt.
\end{aligned}$$

Parce que H' est une densité de probabilité et par les hypothèses (H2) et (H3), nous pouvons écrire :

$$\begin{aligned}
I &\leq C_{x,\theta} \int_{\mathbb{R}} H'(t)\Big(h_k^{b_1} + |t|^{b_2}h_H^{b_2}\Big)dt + F(\theta,y,x) \\
&= \mathcal{O}\Big(h_k^{b_1} + h_H^{b_2}\Big) + F(\theta,y,x)
\end{aligned}$$

Finalement, par hypothèse (H3) nous obtenons

$$\mathcal{J}_2 \longrightarrow F^2(\theta,y,x), \quad si \ \ n \longrightarrow \infty \tag{2.9}$$

La dernière égalité est due au fait que H' est une densité de probabilité, donc nous avons par hypothèse (H3)

$$\int_{\mathbb{R}} H'(t)(F(\theta,y-h_Ht,X_1) - F(\theta,y,x))dt \leq \int_{\mathbb{R}} H'(t)\Big(|t|^{b_2}h_H^{b_2} + h_k^{b_1}\Big)dt \underset{n\longrightarrow\infty}{\longrightarrow} 0$$

✓ **Concernant** J_{2n}

Nous avons par intégration par parties et changement de variables

$$\begin{aligned}
J_{2n} &= \mathbb{E}\Big(H_1(y)| < \theta, X_1 > \Big)\\
&= \mathbb{E}\Big(H\Big(\frac{y-Y_1}{h_H}\Big)| < \theta, X_1 > \Big)\\
&= \int H\Big(\frac{y-v}{h_H}\Big)f(\theta, v, X_1)dv\\
&= \int H\Big(\frac{y-v}{h_H}\Big)dF(\theta, v, X_1)\\
&= \int H'(t)F(\theta, y-h_H t, X_1)dt\\
&= F(\theta, y, x)\int H'(t)dt + \int H'(t)(F(\theta, y-h_H t, x) - F(\theta, y, x))dt,
\end{aligned}$$

la dernière égalité est due au fait que H' est une densité de probabilité. Ainsi, nous avons :

$$J_{2n} = F(\theta, y, x) + \mathcal{O}\big(h_K^{b_1} + h_H^{b_2}\big) \tag{2.10}$$

Enfin, on obtient que $J_{2n} \underset{n\longrightarrow\infty}{\longrightarrow} 0$ Pendant ce temps, par (H1), (H2), (H4) et (H5), il s'ensuit que :

$$\frac{\Phi_{\theta,x}(h_K)\mathbb{E}K_1^2(\theta, x)}{\mathbb{E}^2K_1(\theta, x)} \underset{n\longrightarrow\infty}{\longrightarrow} \frac{C_2(\theta, x)}{(C_1(\theta, x))^2}$$

Puis, en combinant les équations 2.5-2.10, elles conduit à

$$V_n(\theta, y, x) \underset{n\longrightarrow\infty}{\longrightarrow} \frac{C_2(\theta, x)}{(C_1(\theta, x))^2}F(\theta, y, x)(1 - F(\theta, y, x)) \tag{2.11}$$

2.2.2.2 Preuve de lemme 2.2 :

Nous établirons la normalité asymptotique de $A_n(\theta, t, x)$ convenablement normalisée. On a

$$\begin{aligned}
\sqrt{n\Phi_{\theta,x}(h_k)}A_n(\theta, t, x) &= \frac{\sqrt{n\Phi_{\theta,x}(h_k)}}{n\mathbb{E}K_1(\theta, x)}\sum_{i=1}^{n}N_i(\theta, y, x)\\
&= \frac{\sqrt{n\Phi_{\theta,x}(h_k)}}{\sqrt{n}\mathbb{E}K_1(\theta, x)}\sum_{i=1}^{n}N_i(\theta, y, x)
\end{aligned}$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Xi_i(\theta, y, x) = \frac{1}{\sqrt{n}}S_n \tag{2.12}$$

D'abord, on peut écrire,

$$\Xi_i = \frac{\sqrt{\Phi_{\theta,x}(h_k)}}{\mathbb{E}K_1(\theta,x)} N_i$$

Donc

$$Var(\Xi_i) = \frac{\Phi_{\theta,x}(h_k)}{\mathbb{E}^2 K_1(\theta,x)} Var(N_i) = V_n(\theta,y,x).$$

En Note que par 2.11, nous avons $Var(\Xi_i) \longrightarrow V(\theta,y,x)$ lorsque n va à en finity. Évidemment, nous avons

$$\frac{n\Phi_{\theta,x}(h_k)}{V(\theta,y,x)}(A_n(\theta,y,x)) = (nV(\theta,y,x))^{-\frac{1}{2}} S_n$$

Ainsi, la normalité asymptotique de $(nV(\theta,y,x))^{\frac{-1}{2}} S_n$, se déduit les résultats suivants

$$|\mathbb{E}\Big\{ \exp\Big(izn^{-\frac{1}{2}}S_n\Big)\Big\} - \prod_{j=0}^{n}\mathbb{E}\Big\{ \exp\Big(izn^{-\frac{1}{2}}\Xi_n\Big)\Big\}| \longrightarrow 0 \tag{2.13}$$

$$\frac{1}{n}\sum_{j=0}^{n}\mathbb{E}(\Xi_j^2) \longrightarrow V(\theta,y,x) \tag{2.14}$$

$$\frac{1}{n}\sum_{j=0}^{n}\mathbb{E}(\Xi_j^2 \mathbb{1}_{\{|\Xi_j|>\varepsilon\sqrt{nV(\theta,y,x)}\}}) \longrightarrow 0 \;\; \textit{pourtout} \;\; \varepsilon > 0 \tag{2.15}$$

Alors que les équations 2.13 et 2.14 montrent que les Υ_j sont asymptotiquement indépendant, vérifier que la somme de leurs variances tend à $V(\theta,y,x)$. Expression 2.15 est condition de la Lindeberg-Feller pour une somme des termes indépendants.

La normalité asymptotique de S_n est une conséquence des équations 2.13-2.15.

2.2.2.3 Preuve de 2.13 :

Nous utilisons lemme de Volkonskii et Rozanov (Voir l'annexe Masry (2005)) et le fait que le processus (X_i) est i.i.d. En note qu'en utilisant ce $V_j = \exp izn^{-\frac{1}{2}} S_n$, en a

$$\Big|\mathbb{E}\Big\{ \exp\Big(izn^{-\frac{1}{2}}S_n\Big)\Big\} - \prod_{j=0}^{n}\mathbb{E}\Big\{ \exp\Big(izn^{-\frac{1}{2}}\Xi_n\Big)\Big\}| \longrightarrow 0$$

que n tend vers l'infini.

2.2.2.4 Preuve de 2.14 :

En note $Var(S_n) \longrightarrow V(\theta, y, x)$ par l'equation 2.11 et 2.12(par la définition de Ξ_i). Puis que

$$\mathbb{E}(S_n)^2 = Var(S_n) = \sum_{j=0}^{n} Var(\Xi_j)$$

et, en utilisant les mêmes arguments que ceux utilisés précédemment dans la preuve du premier terme de l'équation 2.5, on obtient

$$\frac{1}{n}\sum_{j=0}^{n} \mathbb{E}(\Xi_j^2) = Var(\Xi_j)$$

$Var(\Xi_j) \longrightarrow V(\theta, y, x)$.

2.2.2.5 Preuve de 2.15 :

Rappelons que

$$\Xi_j = \sum_{i=0}^{n} \Upsilon_i$$

Enfin, pour mettre en place 2.15 il suffit de montrer que l'ensemble

$$\{|\Xi_j| > \varepsilon\sqrt{nV(theta, y, x)}\}$$

est négligeable pour n suffisamment large.
À l'aide d'hypothèses (H4) et (H5), en a

$$|\Upsilon_i| \leq C(\Phi(\theta, x)h_K)^{-\frac{1}{2}}$$

donc

$$|\Xi_j| \leq C_n(\Phi(\theta, x)h_K)^{-\frac{1}{2}}$$

qui va à zéro quand n va à l'infini.
Depuis

$$|H_i(y) - F(\theta, y, x)| \leq 1$$

Puis pour n suffisamment large, l'ensemble $\{|\Xi_j| > \varepsilon(nV(\theta, y, x))^{-\frac{1}{2}}\}$ est vide, ceci termine la preuve et, par conséquent, celui de la normalité asymptotique de $((nV(\theta, y, x))^{-\frac{1}{2}})S_n$ et le Lemme (2.2).

2.2.2.6 Preuve de lemme 2.3 :

En a

$$\begin{aligned}\sqrt{n\Phi_{\theta,x}(h_k)}B_n(\theta,y,x) &= \frac{\sqrt{n\Phi_{\theta,x}(h_k)}}{\widehat{F}_D(\theta,x)}\Big\{\mathbb{E}\widehat{F}_N(\theta,y,x) - C_1(\theta,x)F(\theta,y,x)\\ &+ F(\theta,y,x)\Big(C_1(\theta,x) - \mathbb{E}\widehat{F}_D(\theta,x)\Big)\Big\}\end{aligned}$$

Tout d'abord, nous pouvons observer quand $n \longrightarrow \infty$

$$\frac{1}{\Phi_{\theta,x}(h_K)}\mathbb{E}\Big[K^l\Big(\frac{< x - X_i, theta >}{h_k}\Big)\Big] \longrightarrow C_l(\theta,x), \quad pour l = 1,2 \tag{2.16}$$

$$\mathbb{E}[\widehat{F}_D(\theta,x)] \longrightarrow C_1(\theta,x) \tag{2.17}$$

et

$$\mathbb{E}[\widehat{F}_N(\theta,y,x)] \longrightarrow C_1(\theta,x)F(\theta,y,x) \tag{2.18}$$

peut être prouvé de la même manière que dans Ezzahrioui et Ould Said (2008) correspondant à leurs lemmes (21) et (2.2). Alors les preuves de 2.16-2.18 sont omises.
Deuxièmement, en utilisant 2.16, 2.17 et 2.18, en a $n \longrightarrow \infty$

$$\Big\{\mathbb{E}\widehat{F}_N(\theta,y,x) - C_1(\theta,x)(\theta,y,x) + F(\theta,y,x)\Big(C_1(\theta,x) - \mathbb{E}\widehat{F}_D(\theta,x)\Big)\Big\} \quad \longrightarrow \quad 0$$

D'autre part

$$\frac{\sqrt{n\Phi_{\theta,x}(h_k)}}{\widehat{F}_D(\theta,x)} = \frac{\sqrt{n\Phi_{\theta,x}(h_k)}\widehat{F}(\theta,y,x)}{\widehat{F}_D(\theta,x)\widehat{F}(\theta,y,x)} = \frac{\sqrt{n\Phi_{\theta,x}(h_k)}\widehat{F}(\theta,y,x)}{\widehat{F}_N(\theta,y,x)} \tag{2.19}$$

Parce que $K(\cdot)H'(\cdot)$ est continue avec l'appui sur $[0,1]$, puis par hypothèses (H3) et (H4) $\forall m = \inf\limits_{t\in[0,1]} K(t)H'(t)$ tels que

$$\widehat{F}_N(\theta,y,x) \geq \frac{m}{h_H\Phi_{\theta,x}(h_K)}$$

Ce qui donne

$$\frac{n\Phi_{\theta,x}(h_K)}{\widehat{F}_N(\theta,y,x)} \leq \frac{\sqrt{nh_H^2\Phi_{\theta,x}(h_K)^3}}{m}$$

Enfin, la preuve du Lemme (2.3) utilisant (H6), est complétée.

Conclusion

La régression semiparamétrique peut être d'une valeur substantielle dans la solution de problèmes scientifiques complexes.Le monde réel est beaucoup trop compliqué pour que l'esprit humain comprenne en détail. Les modèles de régression semiparamétrique réduisent les ensembles de données complexes aux résumés que nous pouvons comprendre.Correctement appliqué, ils conservent les caractéristiques essentielles des données tout en ignorant les détails sans importance, et donc ils aident la prise de décisions avisée.

Bibliographie

[1] Aït Saidi, A., Ferraty, F., Kassa, R., (2005). Single functional index model for a time series. R. Roumaine Math. Pures et Appl. 50, 321-330.

[2] Aït Saidi, A., Ferraty, F., Kassa, R., Vieu, P., (2008). Cross-validated estimation in the single functional index model. Statistics. 42, 475-494.

[3] A. Ait Saidi, F. Ferraty, R. Kassa, P. Vieu, Cross-validated estimations in the single functional index model, Statistics, 42, (2008a), 475-494.

[4] A. Ait Saidi, F. Ferraty, R. Kassa, P. Vieu, Choix optimal du paramretre fonctionnel dans le modèle à indice fonctionnel simple, C. R. Acad. Sci.Paris, Ser. I, 346, (2008b), 217-220.

[5] A. Ait Saidi, F. Ferraty, R. Kassa, Single functional index model for time series, Romanian J. Pure & Applied Mathematics, 50, (2005), 321-330.

[6] Ai.C(1993).A semiparametric maximum likelihood estimator.Econometrica 65(4), 933-963.

[7] Anderws,D,W,K(1991a).Asymptotic normality of series estimators for nonparametric and semiparametric regression models.Econometrica.59,307-345.

[8] Bachmeier ,L.J.and Q.Li.(2002).Is the term structure nonlinear ? a semiparametric inverstigation.Applied Economitrics Letters 9,151-153.

[9] Box, G. & Cox, D.(1964). An analysis of transformations, Journal of the Royal Statistical Society, Series B 26,211-243.

[10] Cardot, H., Ferraty, F. & Sarda, P. (1999). Functional Linear Model. Statist. & Prob. Letters, 45, 11-22

[11] Chamberlain , G, (1992).Efficiency bounds for semiparametric regression , Econometrica 60, 567-596.

[12] Comte, F., 2004. Kernel deconvolution of stochastic volatility models. J. Time Series Anal. 25,563-582 .

[13] Dabo-Niang, D. & Serge, G. (2010). Functional semiparametric partially linear model with autoregressive errors. J. Multivariate Analysis, Vol 101, **2** 307-315.

[14] Davydov, Yu.A., 1973. Mixing conditions for Markov chains. Theory Probab. Appl. 18,312-3288

[15] E. Masry, Nonparametric regression estimation for dependent functional data : Asymptotic normality. Stoch. Proc. and their Appl, 115, (2005), 155-177.

[16] Fan, J., Truong, Y.K., 1993. Nonparametric regression with errors in variables. Ann. Statist. 21,1900-1925.

[17] Ferraty et al.Peuch, A., Vieu, P., (2003). Modèle à indice fonctionnel simple. C. R. Mathématiques Paris 336, 1025-1028.

[18] Ferré, L, Dauxois, J. & Yao, A. F. (2001). Un modèle semi-paramétrique pour variables aléatoires hilbertiennes (in french). C. R. Acad. Sci. Paris Série. I Math. **333**, No. 10, 947-952.

[19] Francesco, A., 2005. Local likelihood for non-parametric ARCH(1) models. J. Time Series Anal.26,251-278.

[20] F. Ferraty, A. Rabhi, P. Vieu, Conditional quantiles for functional dependent data with application to the climatic El Nin o phenomenon, Sankhy a :The Indian Journal of Statistics, Special Issue on Quantile Regression and Related Methods, 67 (2), (2005), 378-399.

[21] Härdle, W., Tsybakov, A., 1997. Local polynomial estimators of volatility function in nonparametric autoregression. J. Econometrics 81,223-242.

[22] Härdle,W.,H.Liong and J.Gao.(2000).partially linear models.Heiderlberg :phisica-Verlag.

[23] Härdle w.,M.Muller,S.Sperlish and A.Werwatz.(2004).nonparametric and semiparametric models.Berlin : Springer series in statislics.

[24] Hjort, N.L., Glad, I.K., 1995. Nonparametric density estimation with a parametric start. Ann. Statist. 23,882-904.

[25] Hjort, N.L., Jones, M.C., 1996. Locally parametric nonparametric density estimation. Ann. Statist. 24,1619-1647.

[26] Horowitz, J. L. (1988). Semiparametric M-Estimation of Censored Linear Regression Models. In Advances in econometrics : Parametric and robust inference, ed. G. F. Rhodes and T. B. Fomby. Vol. 7 New York : Elsevier pp. 45-83 binary response mode. Econometrica 60, 505-531 model with an unknown transformation of the dependent variable (pp.14-20)

[27] Horowitz,J.L.and W.Härdle.(1996).Drict semiparametric estimation of single-index models with discrete cavariater.jornal of the American Statistical Association 91,1632-1640.

[28] Ichimura,H.(1993).semiparametric least squares (SLS) and weighted SLS estimation of single-index models.Journal of Econometrics 58,71-120.

[29] Ichimura,H.(2000).Asymptotic distribution of nonparametric and semiparametric estimators with data dependent smoothing parameters.unpublished manuscript University College London.

[30] Klien,R.W.and R.W.Spady.(1993).An efficient semiparametric estimator for binary response models.Econometrica.61,387-421

[31] Klimko, L.A., Nelson, P.I., 1978. On conditional least square estimator for stochastic processes. Ann. Statist. 6,629-642.

[32] Maddala, G. S. (1986). Limited-dependent and qualitative variables in econometrics. Cambridge : Cambridge University Press.

[33] Manski, C. F. (1988). Identification of binary response models. Journal of the American Statistical Association 83(403), 729-738.

[34] McCullagh, P. & Nelder, J. A. (1989). Generalized Linear Models, Vol. 37 of Monographs on Statistics and Applied Probability, 2 edn, Chapman and Hall,London.

[35] M.Ezzahrioui, E. Ould-Said, Asymptotic results of a nonparametric conditional quantile estimator for functional time series, Comm. Statist. Theory Methods, 37 (16-17), (2008), 2735-2759.

[36] Nadaraya, E. A. (1964). On estimating regression, Theory of Probability and its Applications 10,186-190.

[37] Naito, K., 2004. Semiparametric density estimation by local L2-fitting. Ann. Statist. 32,1162-1191.

[38] Nelder, J. A. & Wedderburn, R. W. M. (1972). Generalized linear models, Journal of the Royal Statistical Society, Series A 135(3),370-384.

[39] N. Ling, Q. Xu, Asymptotic normality of conditional density estimation in the single index model for functional time series data, Statistics & Probability Letters, 82 (12), (2012), 2235-2243.

[40] Nengxiang Ling, Zhihuan Li, Wenzhi Yang (2014) Conditional Density Estimation in the Single Functional Index Model for a-Mixing Functional Data, Communications in Statistics - Theory and Methods, 43 :3, 441-454, DOI : 10.1080/03610926.2012.664236

[41] Robinson, P. M. (1988). Root-n consistent semiparametric regression. Econometrica 56, 931-954

[42] S. Attaoui, A. Laksaci, E. Ould-Said, A note on the conditional density estimate in the single functional index model, Statist. Probab. Lett, 81(1), (2011), 45-53.

[43] Vieu, P. (1994). Choice of regressors in nonparametric estimation, Computational Statistics & Data Analysis .

[44] Xia, Y., Tong, H., Li, W.K., 1999. On extended partially linear single-index models. Biometrika 86,831-842.