Available online at www.sciencedirect.com
ScienceDirect
C. R. Acad. Sci. Paris, Ser. I 346 (2008) 217–220
COMPTES RENDUS MATHEMATIQUE
http://france.elsevier.com/direct/CRASS1/

Statistique

# Choix optimal du paramètre fonctionnel dans le modèle à indice fonctionnel simple

Ahmed Ait-Saïdi [a], Frédéric Ferraty [b], Rabah Kassa [a], Philippe Vieu [b]

[a] *Laboratoire de mathématiques appliques, université A. Mira, 06000 Béjaia, Algérie*
[b] *Laboratoire de statistique et probabilités, université Paul-Sabatier, 31062 Toulouse, France*

Reçu le 13 février 2007 ; accepté après révision le 22 novembre 2007

Disponible sur Internet le 16 janvier 2008

Présenté par Paul Deheuvels

**Résumé**

Ce papier s'intéresse au choix automatique du paramètre fonctionnel introduit par le modèle à indice fonctionnel lorsqu'on régresse une variable réelle sur une variable fonctionnelle (i.e. à valeurs dans un espace de dimension infinie). Cette Note a pour objectif de montrer qu'une procédure de sélection de type validation croisée conduit à un choix asymptotiquement optimal de l'indice fonctionnel. De plus, on établit que l'indice fonctionnel ainsi choisi est un estimateur consistent. ***Pour citer cet article : A. Ait-Saïdi et al., C. R. Acad. Sci. Paris, Ser. I 346 (2008).***


**Abstract**

**Optimal functional parameter in the single functional index model.** This Note focuses on an automatical choice of the functional parameter introduced throughout the single functional index model when one regresses a real r.v. on a functional r.v. (i.e. valued in an infinite-dimensional space). This Note aims to prove that a cross-validation-based method leads to an asymptotic optimal choice of the functional index. In addition, one states that such a choice for the functional index is a consistent estimator. ***To cite this article: A. Ait-Saïdi et al., C. R. Acad. Sci. Paris, Ser. I 346 (2008).***


## 1. Introduction

Le cadre général de ce travail est la régression d'une variable aléatoire (v.a.) réelle $Y$ sur une v.a. fonctionnelle $X$ (v.a. à valeurs dans un espace de dimension infinie). Ce type de problème a largement été étudié, tant d'un point de vue linéaire (voir la bibliographie dans Ramsay et Silverman, [10]), que d'un point de vue non-paramétrique (voir Ferraty et Vieu, [4]). L'avantage du cadre linéaire est qu'il conduit à l'estimation d'un paramètre fonctionnel qui permet, d'une certaine façon, de visualiser le lien entre $Y$ et $X$. L'inconvénient est que ce gain en « interprétabilité » se fait au prix d'une hypothèse assez restrictive (i.e. linéarité sur l'opérateur de régression), surtout lorsqu'on ne dispose pas d'outil permettant de représenter $Y$ versus $X$. D'un autre côté, si la généralité du modèle non-paramétrique est satisfaisante dans ce cadre fonctionnel, certains peuvent lui reprocher son manque d'éléments permettant de représenter la

*Adresses e-mail :* ferraty@cict.fr (F. Ferraty), vieu@cict.fr (P. Vieu).

1631-073X/$ – see front matter 
doi:10.1016/j.crma.2007.12.013

relation entre la réponse $Y$ et la v.a. fonctionnelle (v.a.f.) $X$. Un bon compromis entre ces deux types de modélisation (paramétrique/non-paramétrique) est obtenu par extension du modèle à indice simple (Härdle et al., [6], Horowitz et Härdle, [7], Hristache et al., [8]) à la situation où on dispose d'une v.a. explicative fonctionnelle. On parle alors de modèle à indice fonctionnel, lequel consiste à supposer que régresser $Y$ sur la v.a.f. $X$ est équivalent à régresser non-paramétriquement $Y$ sur la projection de $X$ sur une direction fonctionnelle $\theta_0$ (inconnue) génériquement appelée « indice fonctionnel ». Ce modèle est moins restrictif que le modèle de régression linéaire fonctionnel tout en offrant la possibilité de visualiser le lien entre $X$ et $Y$ à travers l'indice fonctionnel $\theta_0$ et la fonction lien introduite par un tel modèle. Après avoir introduit la définition d'un estimateur à noyau, nous proposons une méthode d'estimation de l'indice fonctionnel basée sur un critère de validation-croisée (inspirée par Härdle et Marron, [5], et Marron et Härdle, [9]). On montre qu'un tel choix de l'indice fonctionnel est asymptotiquement optimal.

## 2. Le modèle

### 2.1. Modèle

Soit $\{Z_i = (X_i, Y_i)\}_{i=0,1,\dots,n}$ $n+1$ v.a. i.i.d. suivant la loi de $(X, Y)$ à valeurs dans $\mathcal{H} \times \mathbb{R}$, où $\mathcal{H}$ est un espace de Hilbert séparable muni du produit scalaire $\langle \cdot, \cdot \rangle$ et dont la norme associée est notée $\| \cdot \|$. On s'intéresse au modèle à indice fonctionnel suivant :

$$Y_i = r\big(\langle \theta_0, X_i \rangle\big) + \varepsilon_i, \quad \forall i = 1, \dots, n,$$

où la fonction lien $r$ est inconnue (à valeurs dans $\mathbb{R}$), $\theta_0 \in \mathcal{H}$ est inconnu et, pour $i = 1, \dots, n$, $\varepsilon_i$ est tel que $E(\varepsilon_i | X_i) = 0$. Implicitement, ce modèle suppose que $E(Y|X) = E(Y|\langle \theta_0, X \rangle)$. L'indice fonctionnel $\theta_0$ apparaît comme une direction privilégiée de l'espace $\mathcal{H}$ permettant d'extraire la part d'information contenue dans $X$ expliquant $Y$. Ferraty et al. [2] ont montré que ce modèle est identifiable dès qu'on suppose que $r$ est différentiable et $\langle \theta_0, e_1 \rangle = 1$, où $e_1$ est le premier élément d'une base orthonormée de $\mathcal{H}$, hypothèses que que nous considèrerons vérifiées dans tout ce qui suit.

### 2.2. Estimation

$\forall \theta$ fixé, on pose $r_\theta(\cdot) = E(Y|\langle \theta, \cdot \rangle)$ l'opérateur défini sur $\mathcal{H}$ et à valeurs dans $\mathbb{R}$ et soit $\hat{r}_\theta$ l'estimateur de $r_\theta$ introduit par Ferraty et al. [2] :

$$\forall x \in \mathcal{H}, \quad \hat{r}_\theta(x) = \frac{\sum_{i=1}^n Y_i K(h^{-1} |\langle \theta, X_i - x \rangle|)}{\sum_{i=1}^n K(h^{-1} |\langle \theta, X_i - x \rangle|)},$$

où $h = h_n$ est une suite de réels positifs telle que

(H1) $\lim_{n\to\infty} h = 0$, $\lim_{n\to\infty} \frac{\log n}{nh} = 0$, et $C_2 n^{-\tau_2} < h < C_1 n^{-\tau_1}$, avec $0 < \tau_1 < \tau_2 < 1$,

et $K$ est un noyau assymétrique strictement décroissant tel que :

(H2) $\exists c_1, c_2 > 0,\ c_1 1_{[0,1]}(t) < K(t) < c_2 1_{[0,1]}(t),\ \forall t \in \mathbb{R}$.

## 3. Critères de sélection et validation croisée

Il est clair que $\hat{r}_\theta$ dépend du paramètre fonctionnel inconnu $\theta$. L'objectif principal de cette note est donc de proposer un choix pertinent de $\theta$ basé sur le principe de la validation croisée. Dans ce but, à partir d'une fonction de poids $W(\cdot)$ à support compact $\mathcal{C}$, on introduit :

- *Averaged Squared Error* : $ASE(\theta) = n^{-1} \sum_{j=1}^n (r_{\theta_0}(X_j) - \hat{r}_\theta(X_j))^2 W(X_j)$,
- *Integrated Squared Error* : $ISE(\theta) = E[(r_{\theta_0}(X_0) - \hat{r}_\theta(X_0))^2 W(X_0) | Z_1, \dots, Z_n]$,
- *Mean Integrated Squared Error* : $MISE(\theta) = E(ISE(\theta))$,
- *Cross-Validated criterion* : $CV(\theta) = n^{-1} \sum_{j=1}^n (Y_j - \hat{r}_\theta^{-j}(X_j))^2 W(X_j)$, où, pour $j = 1, \dots, n$, $\hat{r}_\theta^{-j}$ est l'estimateur de $r_\theta$ fondé sur l'échantillon auquel on a retiré $(X_j, Y_j)$.

D'un point de vue pratique, on utilise le critère $CV(\cdot)$ pour choisir $\theta$ car les trois premières quantités font apparaître l'opérateur $r_{\theta_0}$ qui est inconnu et on pose $\theta_{CV} = \arg\min_{\theta\in\Theta_n} CV(\theta)$, où $\Theta_n$ est une partie de $\mathcal{H}$ de taille raisonnable au sens de l'hypothèse suivante :

(H3) $\operatorname{card}\Theta_n = [cn^\alpha]$ avec $\alpha > 0$ et $c > 0$.

## 4. Equivalences, optimalité et consistence

Avant d'énoncer les principaux résultats, on considère les hypothèses suivantes :

(H4) $\begin{cases} \forall\theta\in\Theta_n,\ P(|\langle\theta, X_1 - X_2\rangle| < h|X_1) = C_{X_1,\theta}h + \mathrm{o}(h),\ \text{p.s.} \\ \text{avec } 0 < \inf_{\theta\in\Theta_n,\ x\in\mathcal{C}} C_{x,\theta} \leqslant \sup_{\theta\in\Theta_n,\ x\in\mathcal{C}} C_{x,\theta} < \infty. \end{cases}$
(H5) $\exists C > 0,\ \exists\beta > 0,\ \forall(x, y)\in\mathcal{H}^2,\ |r_\theta(x) - r_\theta(y)| \leqslant C|\langle\theta, x - y\rangle|^\beta$,
(H6) $\forall k\in\mathbb{N}^*,\ E(|Y|^k|X) \leqslant C_{k,X} < \infty$, p.s.
(H7) $\exists\eta > 0,\ E(Y^2|X = x) = \sigma(x) \geqslant \eta$, avec $\sigma(\cdot)$ continue.

Les hypothèses (H1)–(H3) et (H5)–(H7) sont standards dans la littérature statistique nonparamétrique. L'hypothèse (H4) est propre à notre contexte mais peu restrictive en ce sens qu'elle est vérifiée (au moins) sous des conditions usuelles d'existence et de positivité d'une densité. Le premier théorème donne l'équivalence asymptotique des différents critères introduits précédemment. Le second résultat permet de dire que $\theta_{CV}$ est asymptotiquement optimal au sens de la *MISE*.

**Théorème 4.1.** *Sous les hypothèses* (H1)–(H7)*, on a* :

$$\sup_{\theta\in\Theta_n} \frac{ASE(\theta)}{MISE(\theta)} \to 1,\ \textit{p.s.},\quad \textit{et}\quad \sup_{\theta\in\Theta_n} \frac{ISE(\theta)}{MISE(\theta)} \to 1,\ \textit{p.s.}$$

**Théorème 4.2.** *Sous les hypothèses* (H1)–(H7)*, on a* : $\frac{MISE(\theta_{CV})}{\min_{\theta\in\Theta_n} MISE(\theta)} \to 1$*, p.s.*

**Corollaire 4.3.** *Si, en plus des hypothèses* (H1)–(H7)*, on suppose que $\Theta_n$ est tel que $\theta_0\in\Theta_n$ et $\forall\theta\in\Theta_n, \forall\epsilon > 0$, $\exists\eta > 0$, $\|\theta - \theta_0\| > \epsilon \Rightarrow E[(r_\theta(X) - r_{\theta_0}(X))^2] > \eta$, alors on a*

$$\|\theta_{CV} - \theta_0\| \to 0,\quad \textit{en probabilité}.$$

**Schémas des preuves.** La démonstration du Théorème 4.1 est basée sur les résultats suivants :

(i) $\min_{\theta\in\Theta_n} MISE(\theta) \geqslant \frac{C}{nh}$,
(ii) $nh\sup_{\theta\in\Theta_n} |ISE(\theta) - MISE(\theta)| \to 0$, p.s.,
(iii) $nh\sup_{\theta\in\Theta_n} |ASE(\theta) - MISE(\theta)| \to 0$, p.s.

(i) provient essentiellement du fait que $\mathrm{var}(\hat{r}_\theta(x)) = Cn^{-1}h^{-1} + \mathrm{o}(n^{-1}h^{-1})$ dont la preuve est donnée dans Ferraty et al. [1]. (ii) et (iii) sont obtenus en adaptant le travail de Marron et Härdle [9] à la situation où on sélectionne un paramètre fonctionnel (i.e. $\theta$) et en utilisant les propriétés de convergence uniforme de $\hat{r}_\theta$ (à $\theta$ fixé) obtenues dans Ferraty et Vieu [3]. Il est clair que (i), (ii) et (iii) conduisent à l'équivalence de l'ASE, ISE et MISE. Le Théorème 4.2 est fondé sur la décomposition

$$CV(\theta) = \widetilde{ASE}(\theta) - 2CT(\theta) + \frac{1}{n}\sum_{j=1}^n \bigl(Y_j - r_{\theta_0}(X_j)\bigr)^2 W(X_j),$$

où

$$\widetilde{ASE}(\theta) = \frac{1}{n}\sum_{j=1}^n \bigl(r_{\theta_0}(X_j) - \hat{r}_\theta^{-j}(X_j)\bigr)^2 W(X_j),$$

$$CT(\theta) = \frac{1}{n}\sum_{j=1}^{n}\big(Y_j - r_{\theta_0}(X_j)\big)\big(\hat{r}_\theta^{-j}(X_j) - r_{\theta_0}(X_j)\big)W(X_j),$$

et sur les 2 résultats suivants :

(iv) $nh\sup_{\theta\in\Theta_n}|ASE(\theta) - \widetilde{ASE}(\theta)| \to 0$, p.s.,
(v) $nh\sup_{\theta\in\Theta_n}|CT(\theta)| \to 0$, p.s.

La preuve de (iv) combine les propriétés de convergence uniforme de $\hat{r}_\theta$ (à $\theta$ fixé) obtenues dans Ferraty et Vieu [3] avec des arguments similaires à ceux employés dans Härdle et Marron [5]. (v) utilise de plus des majorations de moments d'ordre $2k$ de U-statistiques. □

### *4.1. Commentaires*

Grâce au Théorème 4.2, l'optimalité de $\theta_{CV}$ a lieu également au sens de ASE et ISE. Ces équivalences, outre le fait qu'elles servent d'outils pour nos démonstrations, sont aussi utiles dans d'autres contextes (simulations, ...). Sous des conditions analogues à celles du Corollaire 4.3, on pourrait aussi établir la convergence de $\hat{r}_{\theta_{CV}}$ vers $r_{\theta_0}$. Par ailleurs, cette méthode permet de sélectionner une semi-métrique dans une famille (de semi-métriques) indexée par un paramètre fonctionnel ($d_\theta(x,y) = |\langle\theta, x-y\rangle|$) lorsqu'on s'intéresse à des modèles de régression nonparamétrique sur une v.a. fonctionnelle (voir pour plus détails, Ferraty et Vieu, [4]). Il s'agit d'un premier pas important vers la construction de semi-métriques adaptatives dans un contexte de régression nonparamétrique fonctionnelle.

## Remerciements

Les auteurs remercient tous les participants du groupe de travail STAPH « Statistique Fonctionnelle et Opératorielle » de Toulouse pour leurs commentaires et soutiens permanents ainsi que le rapporteur de cette Note pour la pertinence de ses remarques.

## Références

[1] F. Ferraty, A. Mas, P. Vieu, Nonparametric regression on functional data: inference and practical aspects, Aust. N. Z. J. Stat. 49 (3) (2007) 267–286.
[2] F. Ferraty, A. Peuch, P. Vieu, Modèle à indice fonctionnel simple, C. R. Acad. Sci. Paris, Ser. I 336 (2003) 1025–1028.
[3] F. Ferraty, P. Vieu, Nonparametric models for functional data, with applications in regression, time series prediction and curves discrimination, J. Nonparametr. Statist. 16 (2004) 11–127.
[4] F. Ferraty, P. Vieu, Nonparametric Functional Data Analysis: Theory and Practice, Springer Series in Statistics, Springer, New York, 2006.
[5] W. Härdle, J.S. Marron, Optimal bandwidth selection in nonparametric regression function estimation, Ann. Statist. 13 (1985) 1465–1481.
[6] W. Härdle, P. Hall, H. Ichumira, Optimal smoothing in single-index models, Ann. Statist. 21 (1993) 157–178.
[7] J. Horowitz, W. Härdle, Direct semiparametric estimation of single-index models, J. Amer. Statist. Assoc. 91 (1996) 1632–1640.
[8] M. Hristache, A. Juditsky, V. Spokoiny, Direct estimation of the index coefficient in the single-index model, Ann. Statist. 29 (2001) 595–623.
[9] J.S. Marron, W. Härdle, Random approximations to some measures of accuracy in nonparametric curve estimation, J. Multivariate Anal. 20 (1986) 91–113.
[10] J. Ramsay, B.W. Silverman, Functional Data Analysis, second ed., Springer Series in Statistics, Springer, New York, 2005.