INVESTIGACIÓN & DESARROLLO, No. **12**, Vol. **1**: 84 – 93 (2012)
ISSN 2518-4431

ESTIMACIÓN ROBUSTA DE MODELOS ADITIVOS MEDIANTE EL ALGORITMO DE BACKFITTING
ROBUST ESTIMATION FOR ADDITIVE MODELS USING THE BACKFITTING ALGORITHM

Luis P. Yapu Quispe
Universidad Mayor de San Andrés, Bolivia
luis.yapu@gmail.com
(Recibido el 29 de octubre 2012, aceptado para publicación el 19 de diciembre 2012)

RESUMEN

En este trabajo se presenta un método de estimación y simulación de un modelo aditivo a dos variables mediante *splines* robustos, el método general puede ser aplicado con varias variables. El software utilizado para las simulaciones es S+ y se utiliza explícitamente la función *smooth.splineRob* en una implementación del algoritmo de *backfitting*. La función *smooth.splineRob* ha sido escrita en base al trabajo de Cantoni y Ronchetti [3], en el cual se pone énfasis en la selección robusta del parámetro de suavizamiento utilizando una versión robusta del C_p de Mallows, RC_p, y de la validación cruzada, *RCV*. La existencia de datos extremos o no-normales en la parte estocástica de un modelo aditivo puede provocar una mala estimación del parámetro de suavizamiento, lo que tendrá influencia global en la estimación por *splines*. Para la etapa de simulación se realizan las estimaciones por *splines* clásicos y robustos (con estimación robusta del parámetro). La estimación obtenida es muy convincente pero el tiempo de ejecución del programa es relativamente elevado tanto para RC_p y *RCV*, aun cuando, en ciertos casos, con pocas iteraciones robustas se obtienen ya resultados más útiles que la estimación clásica.

ABSTRACT

This paper presents a method of estimation and simulation of an additive model of two variables using robust splines, but the general method can be applied to several variables. The software used for the simulations is S+ and it uses explicitly the smooth.splineRob function in an implementation of the backfitting algorithm. The smooth.splineRob function has been coded based on the work of Cantoni and Ronchetti [3], which emphasizes the robust selection of the smoothing parameter using a robust version of Mallows' C_p, RC_p , and robust cross validation, *RCV*. The existence of outliers or non-normal data in the stochastic part of the additive model may cause a poor estimate of the smoothing parameter that will influence the overall estimate process. For the simulation stage estimations are performed by classical and robust splines (with robust estimation of the parameter). The estimates obtained are very convincing but the execution time of the program is relatively high for both RC_p and *RCV*, even if, in certain cases, few robust iterations are enough to get better results than the classical estimates.

Palabras Clave: Modelos No-Paramétricos, Modelos Aditivos, *Splines* Robustos de Tipo-M, C_p de Mallows Robusto.
Keywords: Nonparametric Models, Additive Models, *M*-type Robust *Splines*, Robust Mallows' C_p.

1. INTRODUCION

En su forma más simple, el punto de vista de penalización de aspereza *(roughness penalty approach)* es un método para relajar las hipótesis del modelo de regresión lineal clásica.

Se considera el modelo,

$$y_i = f(x_i) + \epsilon_i, \quad i = 1, \dots, n, \tag{1}$$

donde $x_1, \dots, x_n$ son puntos de diseño, ϵ_i variables aleatorias independientes con $E[\epsilon_i] = 0$ y $Var(\epsilon_i) = \sigma^2$, $y_1, \dots, y_n$ son las observaciones de la variable de respuesta y f es una función que se desea estimar. Si no se imponen restricciones en la función f, entonces la suma de los cuadrados de los residuos puede ser minimizada a cero escogiendo una función f que interpola los puntos, pero la función obtenida puede perder todas las condiciones de regularidad y oscilar de forma muy abrupta. En particular, el poder predictivo del modelo obtenido es muy reducido (*overfitting*). Imponer condiciones de regularidad mediante polinomios a trozos pero respetando condiciones de derivabilidad en los puntos de separación (*splines* de interpolación), puede ser útil si se sabe que los datos fueron tomados con alta precisión y el fenómeno bajo estudio es conocido como altamente variable. Sin embargo, incluso en estas condiciones, uno puede estar interesado en la variación lenta o tendencia global de los datos y considerar las variaciones locales como ruido aleatorio. Por lo tanto, una aproximación muy cercana a los datos observados no es el único objetivo al ajustar una curva.

DOI: 10.23881/idupbo.012.1-8i

Existen muchas formas de cuantificar la *suavidad*[1] de una función f definida en un intervalo $[a, b]$. Una, que es bastante intuitiva y que se aplica a una función dos veces continuamente diferenciable, es calcular la expresión $\int_a^b \{f''(x)\}^2 dt$. Se pueden mencionar varias razones del por qué esta expresión es una buena medida de *suavidad*, ver por ejemplo Green y Silverman [1].

El punto de vista de penalización de *roughness* consiste en combinar tanto la suma de residuos como la suavidad de la curva de ajuste mediante un parámetro $\lambda > 0$, definiendo la suma de cuadrados penalizada que debe ser minimizada,

$$\min\left\{\sum_{i=1}^{n}\left(\frac{y_i - f(x_i)}{\sigma}\right)^2 + \frac{1}{2}\lambda\int_a^b \left(f''(t)\right)^2 dt\right\}. \tag{2}$$

Un λ pequeño privilegiará el ajuste a los datos observados pudiéndose obtener curvas altamente oscilantes. Si λ es demasiado grande, entonces la curva obtenida podría perder detalles importantes de los datos. El caso límite $\lambda \to \infty$ lleva a la regresión clásica. A pesar del carácter local de los *splines*, queda claro que la estimación del parámetro λ puede tener un impacto global en la estimación y ajuste de la curva.

Huber [2] introdujo los *splines* robustos de tipo-*M*, los cuales son definidos en la Sección 2. Cantoni y Ronchetti [3] argumentan que la estimación de λ debería también basarse al algún criterio robusto y presentan un método de estimación que puede ser visto como una versión robusta de C_p y de validación cruzada. Ideas similares fueron introducidas primeramente en selección robusta de modelos en regresión [4].

Cantoni y Ronchetti [3] indican algunas aplicaciones posibles de la técnica de selección robusta del parámetro de suavizamiento que incluyen modelos más complejos, como los *modelos aditivos y sus generalizaciones*. Estos modelos multivariados pueden utilizar *splines* de suavizamiento como sus bloques dentro del proceso de estimación. Algunos trabajos iniciales sobre el tema son los de Gu [5], [6]. En la Sección 3. se explica cómo es posible incluir la estimación por *splines* robustos dentro del contexto de los modelos aditivos utilizando el algoritmo de *backfitting* que permite el ajuste no lineal de las variables independientes individuales.

Referencias y trabajos relacionados con el tema incluyen, las referencias generales sobre *splines* de suavizamiento de Wahba [7], Härdle [8] y Green y Silverman [1]; la referencia clásica sobre modelos aditivos y generalizaciones de Hastie y Tibshirani [9] y Hastie *et al.* [10], este último es un libro que reúne muchos métodos estadísticos en un marco coherente tomando en cuenta también consideraciones computacionales. Conceptos generales y teoría sobre estadística robusta puede encontrarse en Huber [11], Hampel *et al.* [12] y Huber y Ronchetti [13]. Los *splines* robustos de tipo-M fueron introducidos por Huber [2], su estudio asintótico por Cox [5] y aspectos computacionales por Utreras [14]. Otras aproximaciones al problema de selección robusta del parámetro de suavizamiento son Leung *et al.* [15] y Oh *et al.* [16] y concentran en la validación-cruzada robusta (RCV). Este trabajo está basado en el método propuesto por Cantoni y Ronchetti [3] e implementado en S+ (función smooth.splineRob) en Cantoni [17]. Yuan [18] presenta otra estrategia para robustez mediante *splines* de suavizamiento con cuantiles combinado con un método de validación cruzada aproximada generalizada (GACV). El trabajo de Yuan se acompaña de simulaciones de Monte-Carlo que incluyen un modelo bivariado. Aunque la función de penalización a dos variables podría ser generalizada a un contexto multidimensional, esto implicaría trabajar con integrales múltiples lo que aumenta el costo computacional, comparada con el costo del método de *backfitting* presentado al final de la Sección 3.

Un punto de vista que ha dado muchos resultados en la teoría de *splines* de suavizamiento robustos es mediante la interpretación de pseudo-datos, que permite relacionar asintóticamente la robustez con la teoría clásica de *splines* a partir del trabajo de Cox [19]. Trabajos derivados de este enfoque son, por ejemplo, los artículos de Cantoni y Ronchetti [3], Oh *et al.* [16] y Oh *et al.* [20]; este último con aplicaciones de regresión con wavelets en datos irregularmente espaciados y supresión de ruido en imágenes digitales. En tratamiento de imágenes médicas, el artículo de Lee y Cox [21] compara varios métodos de suavizamiento robustos (no sólo *splines*) implementados en el lenguaje estadístico *R*, con buenos resultados utilizando la validación cruzada absoluta (ACV) para sus datos de espectroscopia.

Aunque dentro de un contexto específico de las ciencias de la salud y la epidemiología, posiblemente el trabajo más próximo a este artículo es el de Alimadad y Salibian-Barrera [22], quienes utilizan también el algoritmo de *backfitting* para detectar aumentos rápidos en el número de casos de enfermedades infecciosas, con el objetivo que estos datos extremos no afecten demasiado estimaciones sobre la mayoría de los datos. Se trata de un trabajo completo que trata aspectos teóricos, simulaciones y aplicaciones a datos reales en epidemiología. Trabajan con validación cruzada robusta (RCV) en R y no aplican el C_p de Mallows robusto (RC_p) de Cantoni [17] que está implementado en S+, el cual es

[1] En este trabajo, se utiliza el término *suavidad* como equivalente al término inglés *smoothness*, similarmente para palabras derivadas. Este término no es sinónimo matemáticamente de la definición de *función suave*, que implica ser continuamente diferenciable un número suficiente de veces.

utilizado en este trabajo. Ambos criterios, RCV y RC_p, se pueden usar independientemente y pueden dar resultados distintos. Un estudio comparativo bastante completo para el caso no robusto en el contexto de la selección de modelos puede encontrarse en [23]. Cada método tiene sus pros y sus contras respecto a sus propiedades estadísticas, asintóticas, hipótesis para aplicarlas y costo computacional. No se encontraron estudios teóricos comparativos en el caso robusto en la literatura, sobre todo en el contexto de suavizamiento combinado con modelos aditivos.

El artículo está organizado como sigue. La Sección 2 introduce el *spline* robusto de tipo *M* y sus propiedades. La Sección 3 presenta el algoritmo de *backfitting* y explica cómo podrían combinarse con los *splines* robustos. La Sección 4 expone resultados de un estudio de simulación que muestra la estabilidad y calidad de los ajustes obtenidos comparados con otros métodos más clásicos. Finalmente la Sección 5 presenta las conclusiones.

2. SPLINES ROBUSTOS DE TIPO-*M*

Se consideran las parejas de puntos de observaciones (x_i, y_i), i=1,...,n. Se quiere ajustar una función f(.) que minimice la expresión:

$$\sum_{i=1}^{n}\left(\frac{y_i - f(x_i)}{\sigma}\right)^2 + \frac{1}{2}\lambda\int_a^b \left(f''(t)\right)^2 dt \tag{2}$$

La solución[2] a este problema está dada por *splines* cúbicos que son polinomios cúbicos a trozos[3] con ciertas condiciones de regularidad en los puntos de salto.

Típicamente, la estimación automática de λ se realiza minimizando el *error predictivo cuadrado medio* (MSPE) mediante validación cruzada (*cross-validation)* o mediante el estadístico C_p de Mallows. La Figura 1 ilustra dos casos de mala estimación del parámetro de suavizamiento (*smoothing parameter*), a la izquierda un caso de *overfitting*, con poco poder predictivo y, a la derecha, una rigidez demasiado elevada del *spline*, que le hace perder características importantes de los datos.

Para controlar la sensibilidad a los puntos extremos (*outliers*), Huber [2] introdujo los *splines* cúbicos de tipo-*M*, los cuales minimizan el criterio,

$$\sum_{i=1}^{n}\sigma\rho\left(\frac{y_i - f(x_i)}{\sigma}\right) + \frac{1}{2}\lambda\int_a^b \left(f''(t)\right)^2 dt \tag{3}$$

Con λ>0, σ es la desviación estándar como en la Ec. (1), y ρ una función convexa. Los detalles técnicos y demostraciones pueden ser encontrados en [3]. Se expondrá aquí solamente los puntos principales.

Se puede escribir la forma de dimensión finita de (3) como:

$$\sum_{i=1}^{n}\sigma\rho(\varepsilon_i) + \frac{1}{2}\lambda f^T K f \tag{4}$$

donde $\varepsilon_i = \frac{y_i - f(x_i)}{\sigma}$, $f = f(x_1,...,x_n)$, $K = N^T\Omega N^{-1}$ y $\Omega_{ij} = \int N_i''(x)N_j''(x)dx$ siendo N una matriz base de *splines*[4].

Definiendo $\psi(t) \equiv \frac{\partial}{\partial t}\rho(t)$ las ecuaciones de estimación son entonces:

$$-\psi(r) + \lambda K\hat{f} = 0 \tag{5}$$

donde $r = (r_1,...,r_n)$, $r_i = \frac{y_i - \hat{f}_i}{\sigma}$, y $\hat{f}$ es la estimación de la función buscada f.

Escogiendo ρ tal que ψ sea acotada, se asegura la robustez respecto a *outliers* en los residuos. Para asegurar la robustez del método completo, se debe estimar σ de manera robusta, el método recomendado es la propuesta 2 de Huber [11] y [12]. Si existen puntos de apoyo (*leverage-points)* en los puntos de diseño x_i es posible realizar el proceso introduciendo antes una matriz diagonal W de pesos, resolviendo $-W\psi(r) + \lambda K\hat{f} = 0$ y obteniendo así un *spline* ponderado.

[2] Dentro de la familia de funciones derivables y con primera derivada absolutamente continua.
[3] Funciones definidas por intervalos y que son polinomios cúbicos en cada intervalo.
[4] N_{ij}=N_j(x_i), con N_j una base de funciones spline.

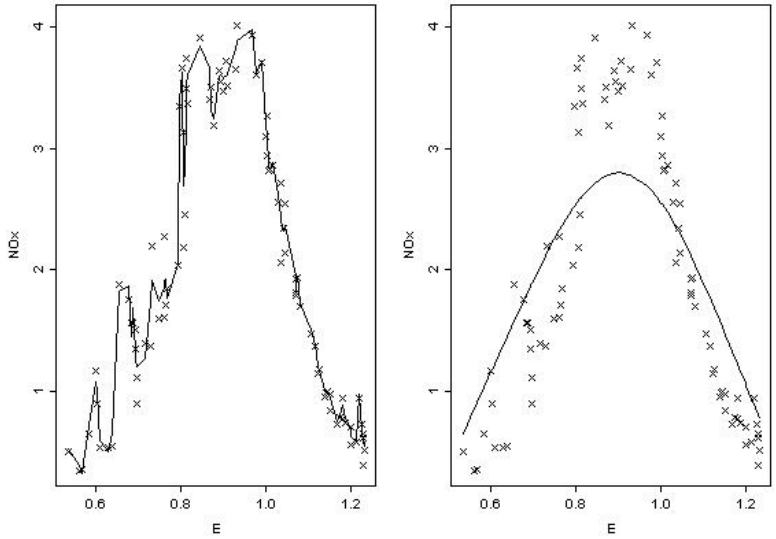

Figura 1 - Base de datos *Ethanol*. Izq: $\lambda = 10^{-10}$, demasiados detalles que podrían ser sólo ruido aleatorio. Der: $\lambda = 0.1$, pierde información importante en la parte central.

Cantoni y Ronchetti [3] presentan un método para la estimación robusta del parámetro de suavizamiento λ. Este método de estimación ha sido programado en la función smooth.splineRob y está incluida en la librería robust de S+. *Cross-validation* y C_p de Mallows son criterios clásicos para predicción. Desde un punto de vista robusto, un criterio no debiera penalizar valores de λ que ajustan bien la mayoría de los datos con la excepción posible de unos pocos puntos (*outliers*).

Sin pérdida de continuidad, en una primera lectura es posible pasar directamente a la Sección 3. Los estimadores de C_p de Mallows robusto, RC_p, y de *Cross-validation* robusto, *RCV*, están dados en las ecuaciones (8) y (9), respectivamente.

Basándose en el trabajo de Ronchetti y Staudte [4], Cantoni y Ronchetti [3] definen el error cuadrado predictivo ponderado,

$$WPSE(\lambda) = \frac{1}{\sigma^2} E\left[\sum_{i=1}^{n} \hat{\omega}_i^2 \left(\hat{f}(x_i) - f(x_i)\right)^2\right] \tag{6}$$

donde $\hat{\omega}_i = \frac{\psi(r_i)}{r_i}$. *WPSE* tiene la forma de un error cuadrado medio ponderado, cuyos pesos tienen el efecto de reducir la contribución $(\hat{f}(x_i) - f(x_i))^2$ de algunos puntos extremos y no penaliza a la mayoría de los datos para los cuales el ajuste funciona bien.

Definiendo la suma ponderada de cuadrados de los residuos por:

$$WSR(\lambda) = \sum_{i=1}^{n} \hat{\omega}_i^2 r_i^2 = \sum_{i=1}^{n} \psi(r_i) \tag{7}$$

y denotando $\delta_i = \frac{\hat{f}(x_i) - f(x_i)}{\sigma}$, Cantoni y Ronchetti proponen la versión robusta del C_p de Mallows dada por:

$$RC_p(\lambda) = WSR(\lambda) - \sum_{i=1}^{n} E\left[\hat{\omega}_i^2 \varepsilon_i^2\right] + 2\sum_{i=1}^{n} E\left[\hat{\omega}_i^2 \varepsilon_i \delta_i\right] \tag{8}$$

La fórmula final en el contexto de *splines* es presentada en la ecuación (12) de [3] y deducida en el apéndice de dicho artículo. En esta fórmula final aparece σ y, de acuerdo a Hastie y Tibshirani [9], es importante tomar una estimación externa de este parámetro que tenga poco suavizamiento. Una estimación robusta de σ está dada como solución de $\sum_{i=0}^{n} \chi\left(\frac{\tilde{r}_i}{\sigma}\right) = 0$, con $\tilde{r}_i = \frac{1}{\sqrt{2}}(y_i - y_{i-1})$, $y_0 = 0$ y $\chi(t) = t\psi(t) - \rho(t) - \beta$, siendo β una constante que asegure la consistencia de Fisher para la estimación de σ. Utreras [14] presenta mayores detalles de los aspectos computacionales

de estos cálculos. De manera alternativa, Cantoni y Ronchetti [3] proponen la siguiente versión robusta del criterio de validación cruzada:

$$RCV(\lambda) = \frac{1}{n}\frac{\sigma^2}{(E\psi')^2}\sum_{i=1}^{n}\frac{\psi^2(r_i)}{(1-S_{ii})^2} \tag{9}$$

donde S_{ii} es el *i*-ésimo término de la diagonal de la matriz $S = \left(I + \frac{\lambda\sigma}{E\psi'}K\right)^{-1}$. Una propuesta similar fue sugerida en Leung *et al*. [24] para suavizadores de kernel de tipo-*M*.

3. MODELOS ADITIVOS Y EL ALGORITMO DE BACKFITTING

Las conclusiones de [3] indican que aplicaciones posibles de la técnica de selección robusta del parámetro de suavizamiento incluyen modelos más complejos, como los modelos aditivos y sus generalizaciones. En esta sección se explica cómo es posible incluir la estimación de *splines* robustos dentro del contexto de los modelos aditivos utilizando el algoritmo de *backfitting*.

Un *modelo aditivo* está dado por:

$$Y = \alpha + \sum_{j=1}^{p} f_j(X_j) + \varepsilon \tag{10}$$

donde los errores $\boldsymbol{\varepsilon}$ son independientes de los X_j, $E[\varepsilon] = 0, Var[\varepsilon] = \sigma^2$ y $E[f_j(X_j)] = 0$. El algoritmo de *backfitting* es un algoritmo muy general que permite el ajuste no lineal utilizando otros métodos de ajuste o regresión para los X_j individuales, sin embargo, no es necesario que los X_j sean de dimensión 1.

El objetivo básico es modelar la dependencia de Y respecto a $X_1,...,X_n$. La herramienta estándar del estadístico aplicado es el modelo clásico de regresión lineal múltiple:

$$Y = \alpha + X_1\beta_1 + ... + X_n\beta_n + \varepsilon \tag{11}$$

Con $E[\varepsilon] = 0, Var[\varepsilon] = \sigma^2$. Este modelo asume que $E[Y]$ depende linealmente de $X_1,...,X_n$, por lo que se hace necesario un solo coeficiente para estimar el f_j, lo que a su vez provee descripciones simples de los datos y métodos sencillos de predicción. Este modelo tiene una característica básica en inferencia estadística: el modelo es *aditivo* en los efectos de los predictores, es decir, la variación de la respuesta estimada sólo depende de un predictor, si se mantienen fijos los otros. Ésta es la característica básica que mantienen los modelos aditivos, además de salvar un problema bastante general de modelos más generales[5], como es el hecho que las vecindades con un número fijo de puntos se vuelven menos locales según aumenta la dimensión (*curse of dimensionality*).

La sección 4.4 de Hastie y Tibshirani [9] expone en orden de generalidad creciente una jerarquía de modelos aditivos, yendo desde la regresión lineal múltiple a la estimación de los f_j con un operador suavizante o suavizador (*smoother*) arbitrario. El algoritmo de *backfitting* es un método general que permite ajustar modelos aditivos utilizando cualquier método de ajuste en sus bloques. En contrapartida, se trata de un método iterativo y éste es el precio que se debe pagar por añadir tal grado de generalidad.

La motivación intuitiva del algoritmo de *backfitting* proviene de las esperanzas condicionales. Si el modelo aditivo (10) es correcto se debe tener $f_j = E\left[y - \alpha - \sum_{k\neq j} f_k(X_k) \mid X_j\right]$. Esto sugiere un método iterativo para estimar los f_j. Para obtener las estimaciones iniciales f_j^0, una buena primera estimación puede ser dada por una regresión robusta de Y en función de los predictores.

De una manera general, el algoritmo de *backfitting* está dado como sigue:

i) Inicializar: $\alpha = promedio(y_j)$, $f_j = f_j^0$, j=1,…,p

ii) Para cada j=1,…,p, actualizar: $f_j = S_j\left(y - \alpha - \sum_{k\neq j} f_k \mid x_j\right)$

iii) Repetir ii) hasta que los f_j se estabilicen.

[5] Por ejemplo los suavizadores de superficie: $Y = f(X_1,...,X_P) + \varepsilon$.

Los S_j pueden ser cualquier operador de ajuste (matrices en la implementación práctica), por ejemplo proyecciones, *splines* o *splines* robustos.

En el capítulo 5 de Hastie y Tibshirani [9] se realiza un estudio teórico del algoritmo y, en particular, se estudian las condiciones para la convergencia de éste. Aquí se explica brevemente la formulación del modelo mediante espacios de Hilbert, ya que es en sí misma matemáticamente interesante y relaciona el algoritmo de *backfitting* con el método de Gauss-Seidel. Lo que sigue, hasta el fin de esta sección, requiere más base matemática y cálculo numérico y puede ser obviada en una primera lectura por el lector más interesado en la parte práctica del método.

Se considera primero el modelo con variables aleatorias (dimensión infinita). Sea $\boldsymbol{\mathcal{H}_j}$ el espacio de Hilbert $\boldsymbol{L_2}$ de funciones medibles $\boldsymbol{\varphi_j(X_j)}$ tales que $\boldsymbol{E[\varphi_j] = 0, E[\varphi_j{}^2] < \infty}$ y producto interior $\boldsymbol{\langle \varphi_j | \varphi_k \rangle := E[\varphi_j \varphi_k]}$. Sean además $\boldsymbol{\mathcal{H}}$ y $\boldsymbol{\mathcal{H}_{XY}}$ los espacio de Hilbert de funciones centradas y cuadrado integrables en $\boldsymbol{p}$ variables: $\boldsymbol{X_1, \dots, X_p}$ y en $\boldsymbol{p+1}$ variables: $\boldsymbol{Y, X_1, \dots, X_p}$, respectivamente. El subespacio $\boldsymbol{\mathcal{H}^{add} := \mathcal{H}_1 + \cdots + \mathcal{H}_p}$, es cerrado bajo condiciones técnicas [25]. Se tiene entonces la jerarquía de subespacios: $\boldsymbol{\mathcal{H}_j \subset \mathcal{H}^{add} \subset \mathcal{H} \subset \mathcal{H}_{XY}}$.

Se desea minimizar entonces $\boldsymbol{E[(Y - g(X))^2]}$ sobre las funciones $\boldsymbol{g(X) := \sum_j f_j(X_j) \in \mathcal{H}^{add}}$. La condición que hace que el problema no sea trivial es la restricción de aditividad. Como $\boldsymbol{\mathcal{H}^{add}}$ es un subespacio cerrado, este mínimo existe y es único. Se denota $\boldsymbol{P_j}$ al operador esperanza condicional $\boldsymbol{E[.|X_j]}$ i.e. la proyección ortogonal sobre $\boldsymbol{\mathcal{H}_j}$. Los residuos $\boldsymbol{Y - g(X)}$ son entonces ortogonales a cada $\boldsymbol{\mathcal{H}_j}$, es decir $\boldsymbol{P_j(Y - g(X)) = 0, j = 1, \dots, p}$.

Para cada variable se tiene:

$$f_j(X_j) = P_j\left(Y - \sum_{k \neq j} f_k(X_k)\right) = E\left[Y - \sum_{k \neq j} f_k(X_k) \mid X_j\right] \tag{12}$$

que se puede escribir como una ecuación matricial cuyos componentes son operadores:

$$Pf = QY \tag{13}$$

donde $f = (f_1(X_1), \dots, f_P(X_P))$ y $Q = diag(P_1, \dots, P_P)$. Entonces, queda explícito el hecho que el algoritmo de *backfitting* es equivalente al método de Gauss-Seidel aplicado al sistema (13).
Esta equivalencia es más clara si se considera el modelo en dimensión finita (versión con datos) y utilizando suavizadores lineales. En efecto, estos suavizadores pueden ser escritos como $\hat{f} = Sy$, siendo S una matriz n x n y n el número de observaciones. Los suavizadores por *spline* caen dentro de este contexto. Volviendo al algoritmo de *backfitting*, a cada operador P_j corresponde una matriz S_j, lo que produce un sistema matricial de tamaño np x np:

$$\hat{P}f = \hat{Q}Y \tag{14}$$

Se observa que los métodos estándares no iterativos para resolver el sistema (14), como la descomposición QR, realizarían $O((np)^3)$ operaciones. Por otra parte, asumiendo que un número constante de iteraciones son realizadas y cada suavizador se aplica en $O(n)$ operaciones (como es el caso de los *splines*), el algoritmo de *backfitting* requeriría sólo $O(np)$ operaciones. Otras consideraciones más sutiles pueden encontrarse en la Sección 5.2 de Hastie y Tibshirani [9].

4. SIMULACION CON DATOS CONTAMINADOS

Se examina la efectividad del método a través de un ejemplo simulado, lo que tiene la ventaja que se conocen las verdaderas funciones f_j de la Ec. (10). Como en el capítulo 9 de Hastie y Tibshirani [9], se generan 100 observaciones a partir del modelo:

$$y = \frac{2}{3}\sin(1.3x_1) - \frac{3}{10}x_2^2 + \varepsilon \tag{15}$$

donde x_1 y x_2 provienen de una ley binormal estándar con correlación 0.4 y $\boldsymbol{\varepsilon}$ de una ley normal contaminada $(1-\omega)N(0,1) + \omega N(0, 10)$ con $\omega = 0.05$ (5% de contaminación).

Se programó el algoritmo de *backfitting*, con p=2 y los operadores S_j implementados ya sea mediante las funciones S+ smooth.spline (para el caso clásico) o smooth.splineRob (para el caso robusto con estimación robusta de λ). Se utilizó una regresión robusta para obtener las funciones iniciales f_j^0.

La primera simulación utiliza el criterio RC_p de Mallows. Se muestra el resultado de la simulación y estimación en la Figura 2, donde además de los puntos simulados (cruces) se estimaron los *splines* clásicos y robustos para cada función del modelo aditivo. La estimación de λ y del *spline* clásico está afectada por los valores extremos y entonces la estimación no captura la "estructura global" dada por la mayoría de los datos, la cual es mejor estimada por el *spline* robusto, línea segmentada de la Figura 2.

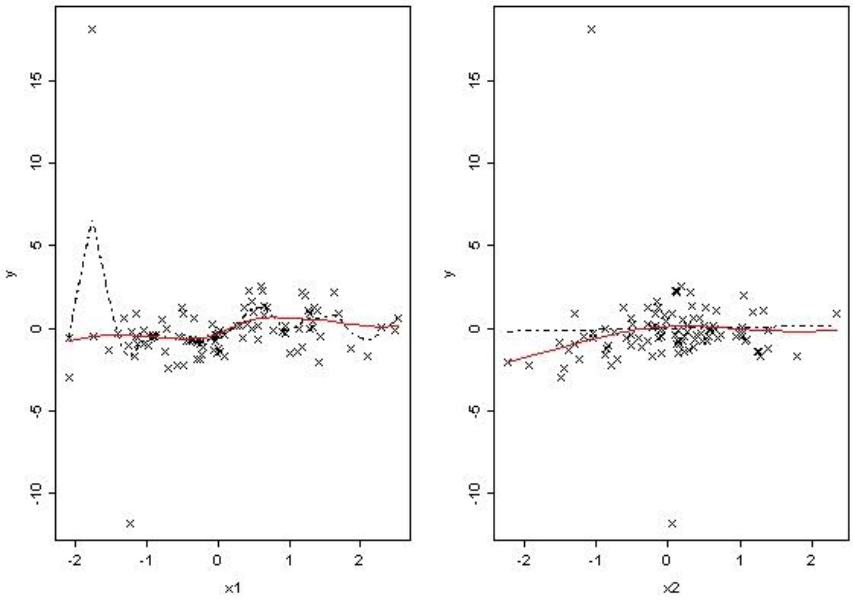

Figura 2 - Izq: Scatterplot para la variable x_1. Der: Scatterplot para la variable x_2. Spline clásico: línea segmentada. Spline robusto (RC_p): línea continua.

El resultado de tres iteraciones del algoritmo de *backfitting* utilizando el criterio de validación cruzada robusta es mostrado en la Figura 3. El *spline* clásico, en línea segmentada, muestra demasiado suavizamiento; el *spline* robusto (con *RCV*), en línea continua, ajusta mejor la verdadera función, Ec. (15). El tiempo ejecución para las tres iteraciones del algoritmo con estimación robusta del parámetro de suavizamiento es del orden de 1 min en una máquina Intel Core i7 1.60 GHrz y 4 GB de RAM, sin utilizar paralelismo.

Considerando el costo computacional, es interesante observar que en todas las pruebas realizadas no se observó diferencia importante en tiempo de ejecución entre RCV y RC_p. Cada llamada a smooth.splineRob toma entre 20-25 segundos para ambos criterios aunque con un tiempo de ejecución en general diferente para los mismos datos. La Figura 3 muestra tres iteraciones consecutivas utilizando el criterio RCV.

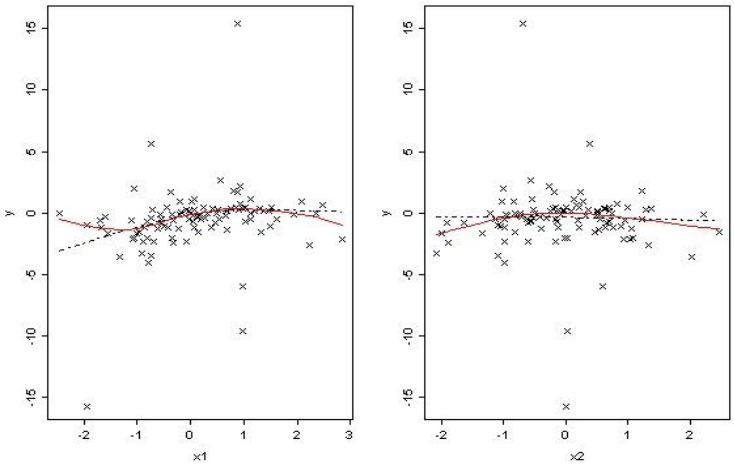

Figura 3 - …

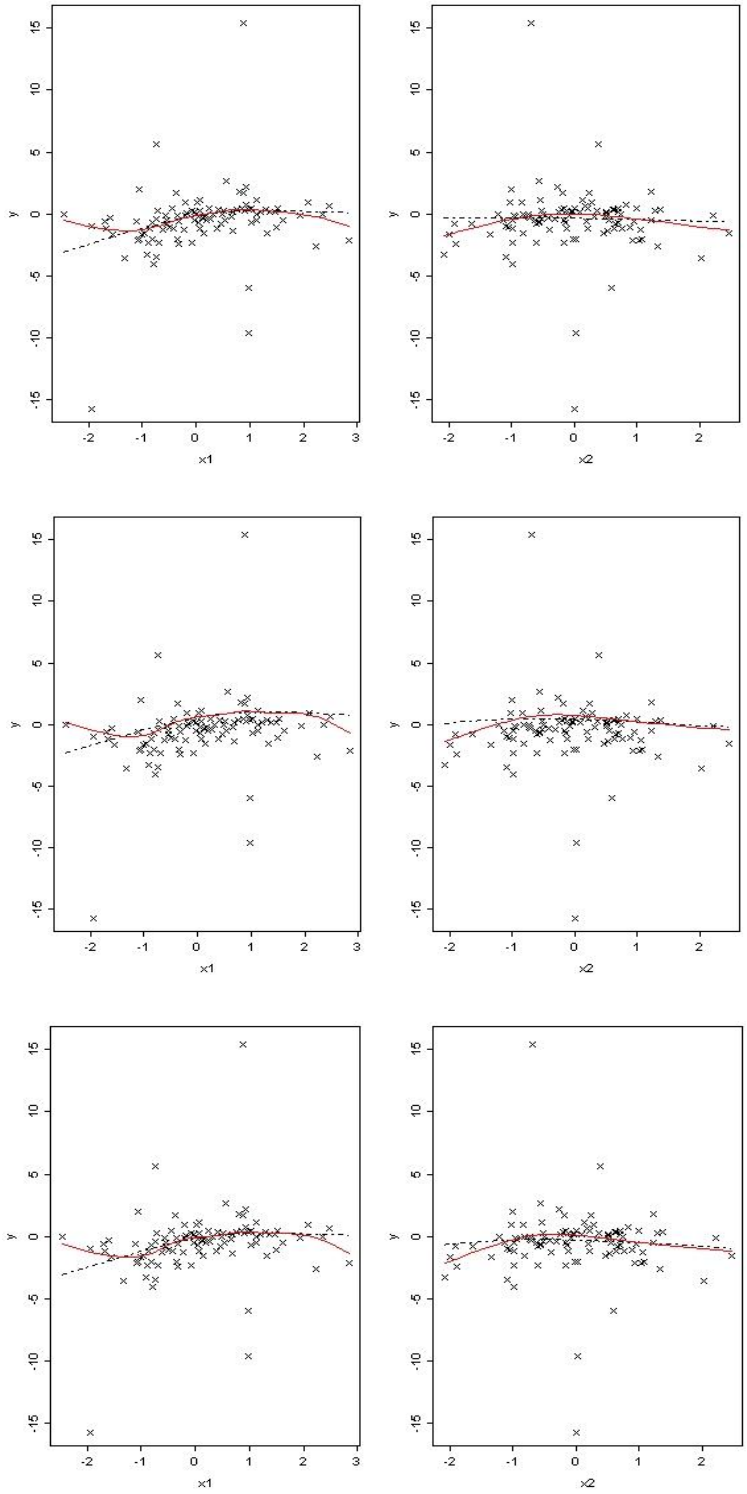

Figura 3 - Izq: Scatterplot para la variable x_1. Der: Scatterplot para la variable x_2. Tres iteraciones del algoritmo de *backfitting*. *Spline* clásico: línea segmentada. *Spline* robusto (*RCV*): línea continua.

Se debe tomar en cuenta que el tiempo de ejecución para la estimación clásica es de menos de 2 segundos para 100 iteraciones, lo que significa que la estimación robusta debe ser todavía optimizada para ser utilizable con los modelos aditivos.

5. CONCLUSIONES

En este trabajo se expuso un método de estimación de un *modelo aditivo*, combinando el algoritmo de *backfitting* con la estimación por *splines* robustos en el cual se pone énfasis en la selección robusta del parámetro de suavizamiento utilizando una versión robusta del C_p de Mallows, RC_p, y de la validación cruzada, RCV. Trabajos anteriores estaban centrados sobre todo en el RCV.

Las simulaciones mostraron que cuando existen *outliers*, la estimación clásica no captura en general el comportamiento global de los datos, y entonces su utilidad es extremamente limitada o nula si uno se interesa en la estructura global de la mayoría de los datos.

La estimación robusta fue muy convincente pero el tiempo de ejecución es relativamente elevado tanto para RC_p como para *RCV*, y no se observaron diferencias significativas en tiempo para ambos criterios. Sin embargo, en la mayoría de las simulaciones, con pocas iteraciones robustas se obtuvieron resultados más útiles que la estimación clásica. La estimación robusta debe ser todavía optimizada para ser utilizable de forma práctica en modelos aditivos multivariados. En particular, se considerará útil implementar el método de *splines* robustos en *R*, el cual se ejecuta más rápido que S+.

6. AGRADECIMIENTOS

El autor agradece al Prof. Elvezio Ronchetti quién propuso el problema y dio pautas importantes para su análisis.

7. REFERENCIAS

[1] P.J. Green y B.W. Silverman. *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*, London: Chapman & Hall, 1994.
[2] P.J. Huber. "Robust Smoothing," in *Launer R.L. y Wilkinson G.N.*, eds. *Robustness in Statistics*, New York/London: Academic Press, 1979, pp. 33-48.
[3] E. Cantoni and E. Ronchetti. "Resistant Selection of the Smoothing Parameter for Smoothing Splines," *Stat. and Comp.* 11, 2001, pp. 141-146.
[4] E. Ronchetti and R.G. Staudte. "A Robust Version of Mallows' Cp," *Journal of the American Statistical Association*, vol. 89, pp. 550-559, 1994.
[5] C. Gu. "Cross-validating Non-Gaussian Data," *Journal of Computational and Graphical Statistics*, vol. 1, pp. 169-179, 1992.
[6] C. Gu. "Diagnostics for Nonparametric Regression Models with Additive Terms," *Journal of the American Statistical Association*, vol. 87, pp. 1051-1058, 1992.
[7] G. Wahba. *Spline models for Observational Data*, Philadelphia: SIAM, 1990.
[8] W. Härdle. *Applied Nonparametric Regression*, Cambridge: Cambridge University Press, 1990.
[9] T.J. Hastie and R. Tibshirani. *Generalized Additive Models*, London: Chapman & Hall, 1990.
[10] T.J. Hastie et al. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2da ed., Springer-Verlag, 2008.
[11] P.J. Huber. *Robust Statistics*, New York: Wiley, 1981.
[12] F.R. Hampel et al. *Robust Statistics: The Approach Based on Influence Functions*, New York: John Wiley, 1986.
[13] P.J. Huber and E. Ronchetti. *Robust Statistics*, 2da ed., New Jersey: Wiley, 2009.
[14] F.I. Utreras. "On Computing Robust Splines and Applications," *SIAM Journal on Scientific and Statistical Computing*, vol. 2, pp. 153-193, 1981.
[15] D.H.Y. Leung. "Cross-validation in Nonparametric Regression with Outliers," *Ann.* Statist, vol. 33, no. 5, 2005, pp. 2291-2310.
[16] H. Oh et al. "Period Analysis of Variable Stars by Robust Smoothing." *J. Roy. Statist. Soc. Ser. C-Applied Statistics*, vol. 53, pp. 15-30, 2004.
[17] E. Cantoni. "Resistant Nonparametric Smoothing with S-PLUS," *Journal of Statistical Software*. [Online]. 10(2), 2004. Avaible: http://www.jstatsoft.org/v10/i02/paper. [Accedido el 26 de diciembre 2012].
[18] M. Yuan. "GACV for Quantile Smoothing Splines," *Comput. Stat. Data An.*, vol. 50, pp. 813-829, 2006.
[19] D.D. Cox. "Asymptotics for M-type Smoothing Splines," *The Annals of Statistics*, vol. 11, pp. 530-551, 1983.
[20] H. Oh et al. "The Role of Pseudo Data for Robust Smoothing with Application to Wavelet Regression," in *Biometrika*, vol. 94, 2007, pp. 893-904.
[21] J-S. Lee and D. Cox, "Robust Smoothing: Smoothing Parameter Selection and Applications to Fluorescence Spectroscopy," in *Computational Statistics & Data Analysis*, vol. 54, no. 12, 2010, pp. 3131-3143.
[22] A. Alimadad and M. Salibian-Barrera. "An Outlier-Robust Fit for Generalized Additive Models With Applications to Disease Outbreak Detection." *Journal of the American Statistical Association*, vol. 106, no. 494, pp. 719-731, 2011.

[23] S. Arlot and A. Celisse. "A Survey of Cross-validation Procedures for Model Selection," in *Statistics Survey*, vol. 4, 2010, pp. 40-79.
[24] D.H.Y. Leung et al. "Bandwidth selection in robust smoothing." *Journal of Nonparametric Statistics*, vol. 2, pp. 333-339, 1993.
[25] I.E. Schochetman et al. "On the Closure of the Sum of Closed Subspaces." *International Journal of Mathematics and Mathematical Sciences*, vol. 26, no. 5, pp. 257-267, 2001.