



**Guillermina Isern**  
**Cristina Cuesta**

*Instituto de Investigaciones Teóricas y Aplicadas en Estadística (IITAE)*

## **ELECCION DEL PARAMETRO DE SUAVIZADO ÓPTIMO EN REGRESIONES P-SPLINE. UN ESTUDIO POR SIMULACION**

### **Introducción**

Las regresiones *spline* Penalizadas (*P-spline*) forman parte del conjunto de técnicas de suavizado que permiten relacionar una variable respuesta con una explicativa sin hacer supuestos a priori de la forma funcional que las liga. En particular, los métodos de regresión *spline* dividen el campo de variación de la variable explicativa en regiones dentro de las cuales se ajusta un modelo de regresión de bajo orden (lineal o cuadrático) que luego son unidos en los puntos extremos (llamados nodos). Por tratarse de un modelo de regresión, es simple estimar la variable respuesta e incluso construir intervalos de confianza y de predicción para dichas estimaciones.

El mayor inconveniente que presentan las regresiones *spline* es que suele ser vulnerable a la cantidad de nodos que se empleen y a la ubicación de los mismos. Una opción para eludir este problema es ponderar los parámetros asociados a los nodos, recurriendo así a los modelos de regresión *P-spline*. Estos últimos modelos dependen del parámetro de suavizado que es quien regula la penalización impuesta a los parámetros. No siempre es simple decidir cuál es el valor del parámetro de suavizado que provoca un mejor ajuste de los datos.

Tradicionalmente se han usado diversos procedimientos para encontrar dicho parámetro: método de validación cruzada (VC), validación cruzada generalizada (VCG), criterio de Mallows ( $C_p$ ), Criterio de información de Akaike (AIC). En la última década se ha propuesto también ajustar las regresiones *P-spline* bajo el enfoque de modelos mixtos lo cual define implícitamente un parámetro de suavizado en función de las componentes de variancia del modelo.

En este trabajo se comparan, mediante un estudio de simulación y bajo diversos escenarios, los parámetros de suavizados encontrados a través de todos estos métodos. El interés se focaliza en comparar el uso de la regresión *P-spline* bajo el contexto de los modelos mixtos con respecto al resto de los métodos ya que bajo esta óptica goza de múltiples ventajas, en principio, no es necesario preocuparse por determinar a priori el parámetro de suavizado y, por otro lado en estos modelos se pueden agregar nuevas covariables al modelo, tener en cuenta estructuras de correlaciones entre las observaciones, etc.

### **Metodología**

Un modelo de regresión *spline* consiste en un ajuste por partes, donde cada una de ellas es una región del campo de variación de la variable explicativa en la que se ajusta un modelo de regresión polinómica (en general de bajo orden) y que están unidos en los extremos



("nodos") para dar continuidad a la curva. En este trabajo se considera una sólo variable regresora, y un polinomio de grado 1, con lo cual la expresión del modelo spline resulta:

$$y_i = f(x_i) + \varepsilon_i = \beta_0 + \beta_1 x_i + \sum_{k=1}^K \beta_{1k} (x_i - N_k)_+ + \varepsilon_i \quad (1)$$

donde las expresiones de la forma  $(x_i - N_k)_+$  se conocen como funciones de "bases truncadas"; las mismas toman el valor  $(x_i - N_k)_+$  si  $x_i > N_k$  y cero en otro caso.

El ajuste resulta más "rugoso" a medida que aumenta la cantidad de nodos y ello implica que puede estar "sobre-ajustando" los datos ya que las pequeñas fluctuaciones que se observan podrían deberse simplemente al azar y no responder a una causa justificada. En oposición, un modelo con pocos nodos puede "sub-ajustar" los datos de manera tal que podría no captar saltos o movimientos que son propios de la verdadera estructura subyacente.

La cantidad de nodos, así como su ubicación podría llevar al analista a probar un gran número de modelos hasta lograr el que le satisfaga y esto implicaría una gran pérdida de tiempo, especialmente si los datos tienen una estructura compleja. Por este motivo la selección de la cantidad y ubicación de los nodos es una problemática que ha sido ampliamente abordada por diversos autores como Ruppert (2003) y Wu (2006) quienes han sugerido que puede resolverse a través del conocimiento profundo de los datos o bien recurriendo a alguna forma de selección automática de los mismos. Un buen ajuste depende en gran medida de la ubicación de los nodos más que de su cantidad. Algunos métodos sugeridos para una correcta ubicación de los nodos son:

- 1) nodos determinados a partir del conocimiento de la estructura subyacente de los datos en términos del problema real (esto es desconocido en la mayoría de las disciplinas),
- 2) nodos determinados por algún criterio de selección del mejor modelo, por ejemplo utilizando un criterio de selección paso a paso (stepwise), este método determina implícitamente, además, la cantidad de nodos a considerar,
- 3) nodos equiespaciados dentro del rango de interés de la variable
- 4) nodos ubicados en percentiles equiespaciados determinados por el conjunto de datos; este es el método utilizado de aquí en adelante en este trabajo

En cuanto a la cantidad de nodos, un número adecuado que se adapta a la mayoría de las situaciones (Ruppert, 2002) es,

$$K = \max \left[ 5, \min \left( \frac{1}{4} \text{Número de valores distintos de } x_i \right), 35 \right]$$

El problema de este tipo de modelos es que dependen de la cantidad nodos que se seleccionen dentro del rango de variación de la variable explicativa y la ubicación de éstos (de las regiones que se consideren y de su amplitud). Como solución a este conflicto, surgen los modelos de regresión Spline penalizados (P-Splines), permitiendo que se defina un número grande de nodos (o regiones) y luego se pondere (o penalice) la importancia de que las regiones sean consideradas distintas.

En efecto, sea el modelo de Regresión Spline (1) con  $K$  nodos, el ajuste por mínimos cuadrados del vector de parámetros  $\beta = (\beta_0, \beta_1, \beta_{11}, \beta_{12}, \dots, \beta_{1K})$  es un vector  $\hat{\beta}$  tal que minimiza  $\|y - X\beta\|^2$ , siendo  $X$  la matriz del modelo donde intervienen las bases truncadas mencionadas. Una forma de lograr un ajuste más "suave" puede llevarse a cabo controlando el peso



de los nodos mediante la imposición de alguna restricción a los parámetros correspondientes a los mismos, por ejemplo  $\sum_{k=1}^K \beta_{1k}^2 < C$ . Esto es equivalente a plantear la minimización de  $\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2$  sujeto a la restricción  $\boldsymbol{\beta}'\mathbf{D}\boldsymbol{\beta} < C$  donde,

$$\mathbf{D} = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}$$

Usando multiplicadores de Lagrange, la función a minimizar es:

$$\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2 + \lambda^2 \boldsymbol{\beta}'\mathbf{D}\boldsymbol{\beta} \quad (2)$$

para algún  $\lambda \geq 0$ . El segundo término de esta expresión es denominado *penalidad de la rugosidad* debido a que es quien produce la rugosidad resultante en la curva ajustada. El grado de suavizado está controlado por  $\lambda$  (al que usualmente se lo denomina *parámetro de suavizado*). La solución de (2) está dada por:

$$\hat{\boldsymbol{\beta}}_{\lambda} = (\mathbf{X}'\mathbf{X} + \lambda^2 \mathbf{D})^{-1} \mathbf{X}'\mathbf{y} \quad (3)$$

El vector de los valores ajustados bajo una Regresión *Spline* Penalizada es entonces:

$$\hat{\mathbf{y}} = \mathbf{X}(\mathbf{X}'\mathbf{X} + \lambda^2 \mathbf{D})^{-1} \mathbf{X}'\mathbf{y} \quad (4)$$

Los coeficientes que son penalizados son los que acompañan a las funciones de base truncada  $(x - N_k)_+$ .

Por lo tanto, en las regresiones *P-Spline* el problema de determinar la cantidad de nodos y su ubicación pierde importancia (ya que su influencia está "controlada" por la penalización), en su lugar, el problema radica en encontrar un valor de  $\lambda$  adecuado (cuyo efecto es muy grande). El caso  $\lambda = 0$  corresponde a un caso de modelo de regresión *Spline* no penalizado. A medida que aumenta  $\lambda$ , el ajuste tiende a ser más suave, al punto que cuando  $\lambda$  es muy grande el efecto de los nodos disminuye significativamente tendiendo a parecerse al ajuste lineal por mínimos cuadrados. En general, si  $\lambda \rightarrow \infty$ , el ajuste tiende a un polinomio del grado utilizado (en este caso, una línea recta).

Se pueden utilizar distintos criterios de selección de dicho parámetro, los que se tuvieron en cuenta en este trabajo son:

- Validación Cruzada (VC):

Una de las medidas más utilizadas de bondad de ajuste de un modelo suele ser la suma de cuadrados residual (*SCR*):

$$SCR = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \|\mathbf{y}_i - \hat{\mathbf{y}}_i\|^2$$

En este caso, el valor ajustado de la variable respuesta depende del parámetro de suavizado  $\lambda$  utilizado en el ajuste del modelo, en efecto:



$$SCR(\lambda) = \sum_{i=1}^n [y_i - \hat{f}(x_i; \lambda)]^2$$

donde  $\hat{f}(x_i; \lambda)$  es el estimador de regresión spline con parámetro de suavizado  $\lambda$ , en el valor  $x_i$ .

Bajo el criterio de "validación cruzada", el modelo que mejor ajusta los datos será aquel que minimicen el valor:

$$VC(\lambda) = \sum_{i=1}^n [y_i - \hat{f}_{-i}(x_i; \lambda)]^2,$$

donde  $\hat{f}_{-i}(x_i; \lambda)$  simboliza el estimador de regresión spline aplicada al conjunto de datos del cual se ha quitado el par  $(x_i; y_i)$ .

- Validación Cruzada Generalizada (VCG)

Sea  $S_\lambda = \mathbf{X}(\mathbf{X}'\mathbf{X} + \lambda^2\mathbf{D})^{-1}\mathbf{X}'$  la matriz de suavizado asociada a  $\hat{f}$ . El vector de valores ajustados es:

$$\begin{bmatrix} \hat{f}(x_1; \lambda) \\ \hat{f}(x_2; \lambda) \\ \dots \\ \hat{f}(x_n; \lambda) \end{bmatrix} = S_\lambda \mathbf{y}.$$

$$\text{Entonces } \hat{f}(x_i; \lambda) = \sum_{j=1}^n S_{\lambda,ij} y_j \quad \text{y} \quad \hat{f}_{-i}(x_i; \lambda) = \frac{\sum_{j \neq i} S_{\lambda,ij} y_j}{\sum_{j \neq i} S_{\lambda,ij}},$$

donde  $S_{\lambda,ij}$  es el elemento  $(i, j)$  de  $S_\lambda$ . Por lo tanto,

$$VC(\lambda) = \sum_{i=1}^n \left( \frac{y_i - \hat{y}_i}{1 - S_{\lambda,ii}} \right)^2.$$

Luego, una aproximación a este valor puede hallarse con:

$$VCG(\lambda) = \sum_{i=1}^n \left( \frac{\{(\mathbf{I} - \mathbf{S}_\lambda)\mathbf{y}\}_i}{1 - n^{-1}tr(\mathbf{S}_\lambda)} \right)^2.$$

Es importante remarcar que VCG no es una generalización de VC, sin embargo, se utiliza por convención esa denominación aunque se podría hablar de "aproximación".

Del mismo modo que en el criterio anterior, el modelo que mejor ajuste será el que minimice esta estadística ( $VCG(\lambda)$ ).

- Criterio  $C_p$  de Mallows



La estadística  $C_p$  de Mallows (1973), en un modelo lineal se calcula como:

$$C_p = SCR(p) + 2\hat{\sigma}_\varepsilon^2 p,$$

donde  $p$  es el número de parámetros del modelo en cuestión.

En el caso de las regresiones *splines* penalizadas, ésta se plantea del siguiente modo:

$$C_p(\lambda) \equiv SCR(\lambda) + 2\hat{\sigma}_\varepsilon^2 gl_{aj}(\lambda),$$

donde  $\hat{\sigma}_\varepsilon^2 = \frac{SCR(\lambda)}{gl_{res}(\lambda)}$ ,  $gl_{res}(\lambda) = n - 2tr(S_\lambda) + tr(S_\lambda S_\lambda^T)$  y  $gl_{aj}(\lambda) = tr(S_\lambda)$ .

El modelo que mejor ajuste tenga será el que minimice esta estadística.

- Criterio de Akaike (AIC)

La estadística a utilizar según este criterio es:

$$AIC(\lambda) \equiv \log[SCR(\lambda)] + \frac{2gl_{aj}(\lambda)}{n}$$

El mejor ajuste lo brindará el modelo que minimice esta estadística.

- Criterio de Akaike Corregido (AICC)

La estadística  $AIC$  es corregida mediante la siguiente modificación:

$$AICC(\lambda) \equiv \log[SCR(\lambda)] + \frac{2[gl_{aj}(\lambda) + 1]}{n - gl_{aj}(\lambda) - 2}$$

Según éste criterio, el mejor ajuste lo brindará un modelo que minimice esta estadística.

- Criterio de Maxima Verosimilitud y Maxima Verosimilitud Restringida

El modelo de regresión *Spline* Penalizada puede expresarse como un modelo mixto considerando  $\beta_{ik}$  como coeficientes aleatorios con distribución Normal con media cero y variancia  $\sigma_\beta^2$ . Esto fue profundizado y difundido en los últimos años por diversos autores (Ruppert et al, 2003; Ngo et al, 2004) y es muy útil tanto por el estudio de sus propiedades teóricas como por su implementación práctica ya que la mayoría de los programas de computación estadísticos tienen rutinas para realizar estimaciones y pruebas de hipótesis con modelos mixtos. Una gran ventaja de este enfoque es que el problema de la cantidad de nodos a utilizar deja de ser importante y además el analista no debe preocuparse por decidir el parámetro de penalidad ( $\lambda$ ) que mejor represente sus datos, ya que el mismo queda determinado por la estimación de las componentes de variancia respectivas ( $\lambda_{opt} = 2m\sqrt{\hat{\sigma}_\varepsilon^2 / \hat{\sigma}_\beta^2}$ , donde  $m$  es el grado del polinomio del ajuste). Si bien existen diversos procedimientos de estimación de



las componentes de variancia, los más razonables están basados en la verosimilitud: Máxima Verosimilitud (MV) o Máxima Verosimilitud Restringida (MVRE).

### Comparación de métodos de selección de $\lambda$ a partir de simulaciones

#### *Modelo de generación de datos*

Los datos fueron generados según el modelo:

$$y = \frac{x^2}{8} + 4 \sin(x) + 4 + \sigma\varepsilon \quad \text{donde } \varepsilon \sim \mathbf{N}(0,1) \text{ y } x \in [-\pi + 4; 4\pi + 4]$$

Se generaron bases de datos de diferentes tamaños muestrales  $n$  (40, 100 y 300) y diferentes parámetros  $\sigma$  (1 y 10). Obteniendo así distintos escenarios definidos por la combinación de tamaño de muestra y parámetro de variabilidad.

En cada escenario se ajustó un modelo de regresión *Spline* de grado uno, con 10 nodos y con distintos parámetros de suavizado que varían entre 0.01 y 10 (con incrementos de 0.01). Para cada escenario se seleccionó un valor  $\lambda$  óptimo de acuerdo a cada uno de los criterios (VC, VCG, AIC, AICC, MV, MVRE).

Este procedimiento se repitió 1000 veces, generando una distribución del  $\lambda$  "óptimo" para cada escenario y cada criterio. En cada una de ellas se calculó la mediana, y los percentiles 25 y 75 (Mna, P<sub>25</sub> y P<sub>75</sub>) estableciendo así una "banda de confianza" para el parámetro de suavizado.



## Resultados

Los resultados provenientes de las simulaciones se muestran en las Tablas 1 y 2. En las mismas se observan las medianas y cuartiles asociados a las distribuciones de los parámetros de suavizado ( $\lambda$ ) bajo los distintos escenarios planteados.

Los principales hechos a destacar son:

- A mayor variancia, mayor  $\lambda$ . Esto significa que a mayor variabilidad en la nube de puntos, es más difícil captar "saltos" intrínsecos y por lo tanto la curva resultante es más suave. Por otro lado, las comparaciones entre métodos de estimación y entre tamaños de muestra son similares dentro de cada una de las variabilidades ( $\sigma$ ), por lo que las siguientes conclusiones se dan tanto para  $\sigma = 1$  como para  $\sigma = 10$ .
- Cuando el tamaño de muestra es mayor, se dispone de más información. Esto redundando en la construcción de curvas más suaves y por lo tanto los parámetros de suavizado son mayores particularmente al utilizar los métodos de estimación como VC VCG CP AIC AICC. En otras palabras, si los tamaños de muestra son pequeños la curva depende mucho de esos datos, en tanto que si el tamaño de muestra es mayor, la curva es más estable (con menos "picos").
- Al comparar los métodos de suavizado, se observa que con los métodos de VC y VCG los valores de  $\lambda$  seleccionados son similares y la variabilidad que presentan también. Siendo un poco mayor el valor de  $\lambda$  que se obtiene con VC respecto al de VCG. Ver Gráfico 1.
- Si se comparan los resultados obtenidos con los métodos de MV y MVRE, se observa que cuando el tamaño de muestra es chico el método de MV tiende a seleccionar valores de  $\lambda$  muy grandes ( $\lambda \rightarrow +\infty$ ) lo que significa que el ajuste que maximiza la verosimilitud de la muestra es el de una regresión simple lineal. En cambio con el de MVR los valores de  $\lambda$  son similares a los que proporcionan los demás métodos, lo que permite realizar un ajuste más acorde al modelo teórico. Ver Gráfico 2.

Por otro lado, para tamaños de muestra mayores las estimaciones de  $\lambda$  son similares en todos los escenarios. En particular se destaca que ambos métodos no se diferencian.

Una explicación de ello puede deberse a las estimaciones de las componentes de variancia provenientes del modelo mixto que se ven representadas en el Gráfico 3. En efecto, cuando el tamaño de muestra es chico, pueden encontrarse componentes del error altas en comparación con las del efecto aleatorio, de lo que resultan valores de  $\lambda$  muy grandes. En cambio esta situación no ocurre cuando el tamaño de muestra es grande.



Tabla N°1: Medianas, percentiles 25 y 75 del parámetro de suavizado  $\lambda$ , cuando en los datos generados se utilizó  $\sigma = 1$ .

Método de Estimación de $\lambda$	Tamaño de Muestra		
	40	100	300
	Mna(P <sub>25</sub> ;P <sub>75</sub> )	Mna(P <sub>25</sub> ;P <sub>75</sub> )	Mna(P <sub>25</sub> ;P <sub>75</sub> )
<b>VC</b>	0.49 (0.38;0.60)	0.57 (0.49;0.65)	0.67 (0.61;0.73)
<b>VCG</b>	0.49 (0.40;0.59)	0.57 (0.49;0.64)	0.66 (0.60;0.72)
<b>CP</b>	0.43 (0.34;0.53)	0.54 (0.46;0.61)	0.65 (0.59;0.71)
<b>AIC</b>	0.42 (0.32;0.51)	0.54 (0.46;0.61)	0.65 (0.59;0.71)
<b>AICC</b>	0.61 (0.52;0.72)	0.61 (0.53;0.68)	0.67 (0.62;0.73)
<b>MV</b>	1,04.10 <sup>10</sup> (0.58;1,89.10 <sup>10</sup> )	0.49 (0.45;0.54)	0.50 (0.47;0.53)
<b>MVRE</b>	0.51 (0.43;0.61)	0.49 (0.45;0.53)	0.50 (0.47;0.52)

Tabla N°2: Medianas, percentiles 25 y 75 del parámetro de suavizado  $\lambda$ , cuando en los datos generados se utilizó  $\sigma = 10$ .

Método de Estimación de $\lambda$	Tamaño de Muestra		
	40	100	300
	Mna(P <sub>25</sub> ;P <sub>75</sub> )	Mna(P <sub>25</sub> ;P <sub>75</sub> )	Mna(P <sub>25</sub> ;P <sub>75</sub> )
<b>VC</b>	0.550 (0.42;0.69)	0.61 (0.51;0.69)	0.69 (0.62;0.75)
<b>VCG</b>	0.55 (0.42;0.68)	0.60 (0.52;0.69)	0.68 (0.62;0.74)
<b>CP</b>	0.48 (0.36;0.60)	0.57 (0.49;0.66)	0.67 (0.61;0.73)
<b>AIC</b>	0.46 (0.34;0.58)	0.57 (0.49;0.66)	0.67 (0.61;0.73)
<b>AICC</b>	0.69 (0.55;0.83)	0.64 (0.56;0.74)	0.70 (0.63;0.75)
<b>MV</b>	1,1.10 <sup>10</sup> (0.78; 1,8.10 <sup>10</sup> )	0.54 (0.49;0.61)	0.52(0.49;0.55)
<b>MVRE</b>	0.61 (0.49;0.79)	0.532 (0.48;0.59)	0.52 (0.49;0.55)





Gráfico 1: Intervalos formados por los  $P_{25}$  -  $P_{75}$  y Mediana de la distribución de  $\lambda$  para todos los criterios utilizados, en los escenarios donde  $n=100$ , para  $\sigma=1$  y  $\sigma=10$

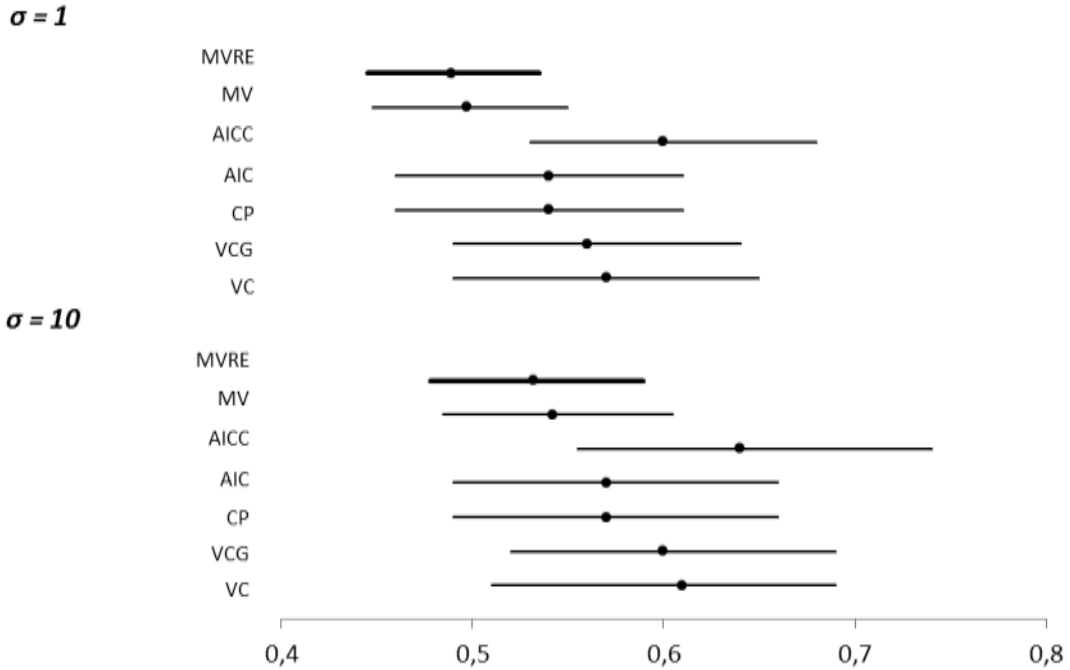


Gráfico 2: Intervalos formados por los  $P_{25}$  -  $P_{75}$  y Mediana de la distribución de  $\lambda$  para los criterios MV y MVRE, en los escenarios donde  $n=100$ ,  $n=40$ , para  $\sigma=1$  y  $\sigma=10$

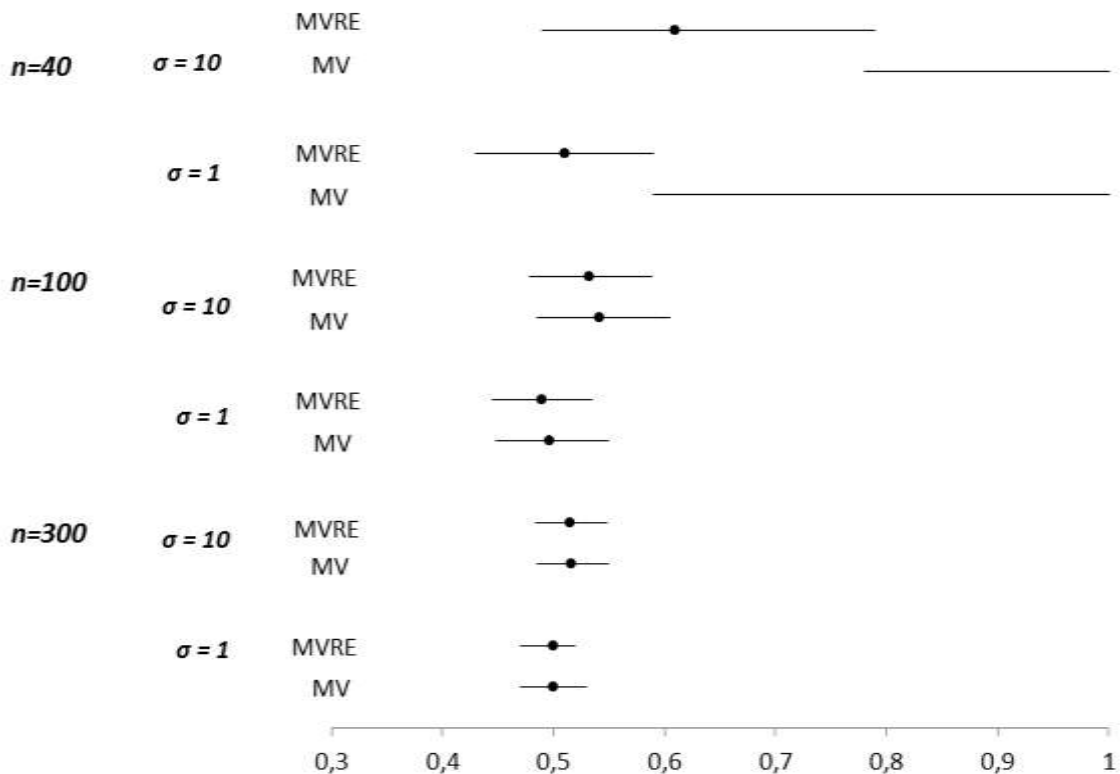
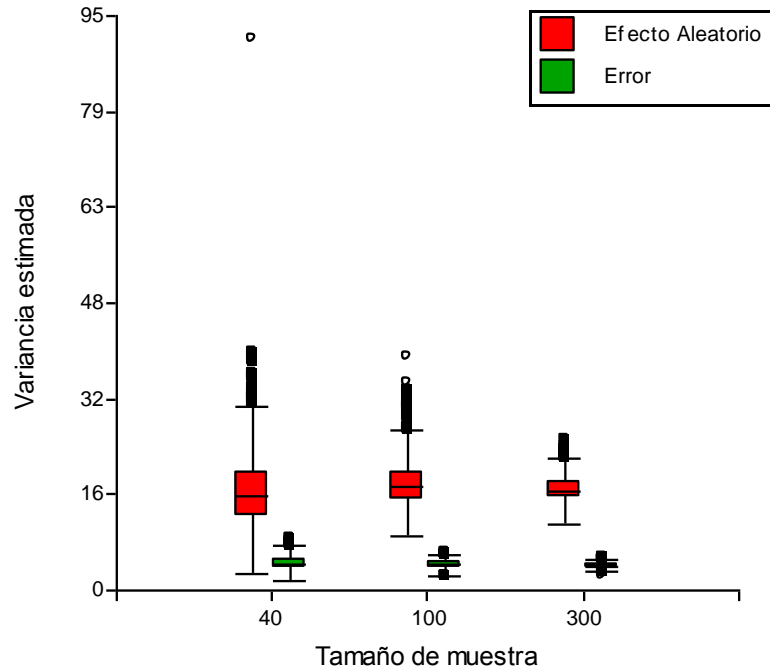




Gráfico 3: Variancias estimadas del efecto aleatorio y del error, según el tamaño de muestra, cuando  $\sigma = 1$



### Discusión

Cuando se llevan a cabo regresiones semiparamétricas tipo *Spline* penalizadas, la principal preocupación es elegir correctamente el parámetro de suavizado. Para ello se han propuesto diferentes métodos. En este trabajo se comparan estos métodos, mediante un estudio de simulación y bajo diversos escenarios.

Los resultados hallados muestran que no hay diferencias notorias entre ellos, por lo cual se recomienda el uso de la regresión *P-Spline* bajo el contexto de los modelos mixtos ya que de esta forma, no es necesario preocuparse por determinar a priori el parámetro de suavizado y, por otro lado estos modelos gozan de otras ventajas, por ejemplo se pueden agregar nuevas covariables al modelo, tener en cuenta estructuras de correlaciones entre las observaciones, etc.

En particular, se recomienda la estimación por MVRE que es la que permanece más estable a los cambios en tamaño de muestra.

Se sugiere continuar con el estudio de esta problemática planteando nuevos escenarios y utilizando el error cuadrático medio como una medida de bondad de ajuste para la comparación.



## Referencias Bibliográficas

- Bowman, A. and Azzalini, A. (1997). *Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustration*. Oxford, UK: University Press
- Durbán, M. (2009). An introduction to smoothing with penalties: P-Splines. *Boletín de Estadística e Investigación Operativa*, Vol. 25, No. 3, pp. 195-205.
- Eubank, R. (1999). *Nonparametric Regression and Spline Smoothing*. New York. Marcel Dekker.
- Fox, J. (2000). *Nonparametric Simple Regression: Smoothing Scatterplots*. Thousand Oaks, CA: Sage
- Gurrin L., Scurrah K. Hazelton M. (2005). Tutorial in biostatistics : Spline Smoothing with Linear Mixed Models. *Statistics in Medicine* 24:3361-3381
- Mallows, C. L. (1973). Some comments on *Cp*. *Technometrics* 15: 661–75.
- McCulloch, C. E.; Searle S. R. (2001). *Generalized, Linear and Mixed models*. New York: Wiley
- Ngo L., Wand M. (2004). Smoothing with Mixed Models Software. *Journal of Statistical Software* Volume 09, Issue 01. URL: <http://www.jstatsoft.org/>
- Ruppert D., (2002). Selecting the Number of Knots for Penalized Splines. *Journal of Computational and Graphical Statistics*. Volume 11(4):735-757
- Ruppert D., Wand M.P., Carroll R. (2003). *Semiparametric Regression*. Cambridge University Press
- SAS Institute Inc. (2009). *SAS/STAT® 9.2 User's Guide: Mixed Modeling (Book Excerpt)*. Cary, NC (USA).
- Simonoff, J.S. (1996). *Smoothing Methods in Statistics*. New York. Springer -Verlag
- Smith P. (1979). Spline as a Useful and Convenient Statistical Tool. *The American Statistician* Volume 33, No 2.
- Wu H., Zhang J. (2006). *Nonparametric Regression. Methods for Longitudinal Data Analysis*. New York: Wiley