UNIVERSIDAD DE BUENOS AIRES

Facultad de Ciencias Exactas y Naturales

Departamento de Matemática

Modelo Lineal funcional con restricciones de forma

Tesis presentada para optar al título de Doctor de la Universidad de Buenos Aires en el área Ciencias Matemáticas

Manuel Eduardo Benjamín

Directora de tesis: Daniela Rodriguez.

Consejera de estudios: Graciela Boente.

Lugar de trabajo: IMAS - Conicet, Facultad de Ciencias Exactas y Naturales - UBA

Buenos Aires, 2020

Modelo lineal funcional con restricciones de forma

Resumen

En este trabajo estudiamos el modelo lineal con covariables funcionales y respuesta escalar cuando existen restricciones sobre la forma de la función de peso. Algunas de las restricciones consideradas son positividad, monotonía, convexidad o de nulidad a partir de un cierto punto.

Estudiamos y desarrollamos herramientas teóricas para obtener resultados de convergencias de elementos aleatorios en un espacio de Hilbert separable. Utilizamos una (semi) norma definida por el operador de covarianza de las covariables y relacionada con el error cuadrático medio de predicción. Demostramos una Ley de los Grandes Números Uniforme para su versión empírica y obtenemos un resultado que vincula las tasas de convergencia en (semi) norma empírica con las de la (semi) norma de interés.

Estos resultados nos permiten obtener, con hipótesis menos restrictivas a las usuales en la literatura de datos funcionales, tasas de convergencia para una amplia familia de estimadores en el modelo sin restricciones y además, demostrar su consistencia bajo la norma inducida por el producto interno del espacio.

Para el modelo restringido proponemos una familia de estimadores que cumplen con las restricciones de forma y mostramos, bajo ciertas condiciones, que sus tasas de convergencia son menores o iguales a las obtenidas para los estimadores sin restricciones. Realizamos un estudio de simulación para evaluar su desempeño en muestras finitas.

Finalmente, consideramos el caso donde la función de peso es monótona y nula a partir de un momento. Proponemos un estimador para el punto de cambio, es decir, el momento a partir del cual la función es idénticamente cero. Probamos su consistencia y mediante un estudio de simulación mostramos su desempeño en distintos tamaños de muestras.

Palabras clave: Modelo lineal funcional, Estimación restringida, Penalización, Datos funcionales, Modelo funcional histórico. Detección de punto de cambio.

Functional linear model with shape constraints

Abstract

In this thesis we study the linear model with functional covariates and scalar response under shape constraints on the weight function. These constraints may be, among others, positivity, monotonicity, convexity or beeing constantly zero after a certain point.

We study and develop theoretical tools for the convergence of random elements in a separable Hilbert space. We use a (semi) norm based on the covariance operator of the functional covariates and related to the mean squared prediction error. We demonstrate an Uniform Law of Large Numbers for its empirical version and obtain a result that links the convergence rate of the empirical (semi) norm to the (semi) norm of interest.

These results allow us to obtain convergence rates for a wide family of estimators in the unconstrained model and show their consistency under the norm induced by the inner product. These results are obtained with milder hypothesis than those commonly used in the functional data literature.

For the constrained model we propose a family of estimators that satisfy the shape constraints. Under certain conditions, we prove that their convergence rates are as good as or better than those obtained for the uncontstrained estimators. We perform a simulation study to compare the prediction errors for finite samples.

Finally we consider a special kind of constraint where the weight function is decreasing and equals zero after a certain point. We define an estimator for the change point, i.e. the first moment where the function is constantly zero. We prove the consistency of the proposed estimator and we conduct a simulation study in order to show its behaviour on different sample sizes.

Palabras clave: Functional linear model, Constrained estimation, Penalization, Functional data. Historical Functional linear Model, Change point detection.

Índice

1. Introducción **6**

1.1. El Modelo Lineal Funcional con respuesta escalar 6

1.2. Un ejemplo de aplicación del Modelo Lineal Funcional 7

1.3. Estimación bajo restricciones 9

1.4. Procesos estocásticos 10

1.5. Procesos aleatorios sobre un espacio de Hilbert 12

1.5.1. Clasificación de operadores lineales 12

1.5.2. Elementos aleatorios en un espacio de Hilbert separable 14

1.5.3. El proceso aleatorio como dato funcional 15

2. Convergencia de elementos aleatorios en un espacio de Hilbert **16**

2.1. Ley de los Grandes Números Uniforme 17

2.2. Relaciones entre distintos tipos de convergencia. 21

3. Modelo lineal con covariable funcional y respuesta escalar **24**

3.1. Modelos de regresión generales 25

3.2. Modelo de Regresión Lineal con restricciones 27

3.2.1. Definición de los estimadores 27

3.2.2. Cálculo de los estimadores 29

3.3. Propiedades para tamaño de muestra fijo 29

3.3.1. Estimador de cuadrados mínimos penalizados 30

3.3.2. Estimador restringido de cuadrados mínimos penalizados 32

3.4. Comportamiento asintótico de los estimadores 32

3.5. Estimación mediante espacios de splines 34

3.5.1. Computo del estimador restringido 35

3.5.2. Resultados asintóticos para la estimación sin restricciones 35

3.5.3. Resultados asintóticos para la estimación restringida 37

4. Modelo Lineal Histórico **39**

4.1. Estimación en dos pasos del punto de cambio 39

4.2. Estimación del modelo con un punto de truncado 40

4.3. Resultados Asintóticos 42
4.3.1. Estimador del punto de cambio 42
4.3.2. Estimación en tres pasos 43

5. Estudio de simulación **46**
5.1. Estimación de la función de peso bajo restricciones de forma 48
5.1.1. Algunas observaciones sobre los errores cuantificados por la Γ-Norma 50
5.1.2. Algunas observaciones sobre los errores cuantificados por la norma L_2 51
5.1.3. Estimación con restricciones cuando no se cumplen las condiciones 52
5.1.4. Estimación Monótona vs estimación Convexa 52
5.2. Estimación en el Modelo Lineal Histórico 52
5.2.1. Estimación del punto de corte 53
5.2.2. Estimación en tres pasos de la función de regresión 54
5.3. Conclusiones 54

6. Ejemplo de aplicación: Emisión de partículas sólidas **55**

7. Demostraciones del Capítulo 1 **64**

8. Demostraciones del Capítulo 2 **66**

9. Demostraciones del Capítulo 3 **72**

10. Demostraciones del Capítulo 4 **82**

11. Apendice **91**
11.1. Propiedades de la traza 91
11.2. Figuras y tablas. 93

12. Bibliografia **146**

1. Introducción

1.1. El Modelo Lineal Funcional con respuesta escalar

El estudio de metodologías estadísticas con datos que se representan como funciones o curvas es un área conocida como Datos Funcionales. Este área ha ganado popularidad desde los finales de la década de 1990 y hoy en día es un tópico en creciente desarrollo. Los datos funcionales pueden ser pensados como curvas o funciones sobre un dominio continuo. Algunos ejemplos usuales son curvas de altura de personas a lo largo del tiempo, espectrometrías, de precios diarios de activos financieros o de temperatura a lo largo del tiempo entre otras.

El modelo de regresión lineal con covariable funcional y respuesta escalar es un pilar fundamental de los modelos de regresión funcionales. En este existe un proceso aleatorio que define una covariable funcional x que vive en $L_2(I)$ con $I \subseteq \mathbb{R}$. Este elemento aleatorio tiene una respuesta escalar asociada que sigue la siguiente ecuación

$$Y = \alpha + \int_I b(t)x(t) + \varepsilon. \tag{1}$$

La función b se conoce como función de peso o función de regresión del modelo. Tanto esta como el término independiente o intercept $\alpha \in \mathbb{R}$ son elementos no aleatorios. El término del error ε es una variable aleatoria con media cero y varianza finita. Este modelo ha sido extensamente estudiado cuando I es un intervalo cerrado y acotado de $\mathbb{R}$. Las propuestas de la literatura se diferencian cuando el interés radica en conocer la función b o el funcional $\Phi(x) = \int_I b(t)x(t)\,dt$. El segundo caso está relacionado con la predicción de la respuesta para un nuevo elemento.

En la mayoría de los casos prácticos los datos funcionales son observados sobre una grilla $t_1, \ldots, t_p$ del intervalo I. Es decir, para cada curva, se cuenta con los valores $(t_j, x(t_j))$ para $1 \leq j \leq p$. Las grillas pueden ser diferentes para cada curva, y se suelen asumir hipótesis de suavidad sobre estas que permiten pasar de los datos grillados al continuo. Las grillas también pueden ser irregulares, es decir, la distancia entre t_j y t_{j+1} no se mantiene constante. Esta situación no resulta un problema en la práctica siempre que la grilla sea lo suficientemente fina, es decir, las distancias entre los puntos de la grilla es suficientemente chica. Los casos en que esto no ocurre pueden presentar problemas metodológicos y teóricos. Este tipo de datos son conocidos como datos funcionales esparsos.

En este trabajo asumimos que contamos con el dato funcional completo, es decir, la curva $x(t)$, $t \in I$. Si bien esta situación no es frecuente en la práctica, si se cuenta con una grilla muy fina o, mejor aún, si se cuenta con la posibilidad de obtener una grilla tan fina como se desee podemos utilizar hipótesis de suavidad sobre las curvas para obtener la curva continua. En muchas situaciones el error que se comete es despreciable ante la variabilidad que presenta el modelo.

Una de las ventajas de los modelos funcionales es que permiten superar el problema de alta dimensión de manera natural en el caso en que se pueda suponer que los datos tienen una estructura continua. Esto marca una importante diferencia de los modelos multivariados usuales. El siguiente ejemplo ilustra esta afirmación. ´

1.2. Un ejemplo de aplicación del Modelo Lineal Funcional

En este ejemplo hacemos uso de los datos de Janssens (1998) que fueron estudiados por Maronna (2013). Los datos se componen de $n = 180$ curvas espectométricas de vasijas de vidrio arqueológicas. El rango de las frecuencias de las espectrometrías fue tomado de 100 a 400 y los datos presentados en una grilla regular con $p = 301$. A cada vasija se le realizó una medición química del nivel de óxido que corresponde a la respuesta Y del modelo 1. Los valores de la grilla los convertimos a datos funcionales que representamos en un espacio de splines S de grilla regular. La Figura 1 muestra algunas curvas y en rojo la curva media. Separamos al azar los datos en $\frac{2}{3}$ para el grupo de entrenamiento y $\frac{1}{3}$ para el grupo de testeo.

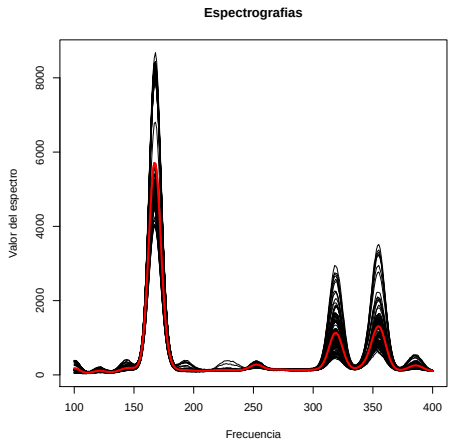

Figura 1: Curvas espectrométricas.

Para ajustar el modelo funcional utilizamos un espacio H de splines de orden cuatro y dimensión 120 sobre una grilla regular. La función de peso del modelo 1 la obtenemos como el elemento de H que minimiza sobre el grupo de entrenamiento la expresión

$$\sum_{i=1}^{120} \left(Y_i - \bar{Y} - \int_{100}^{400} f(t)(x_i(t) - \overline{x}(t))\, dt \right)^2 + \rho \int_{100}^{400} (f''(t))^2\, dt.$$

La integral de la segunda derivada es un término de penalización que ayuda a obtener soluciones más suaves. El valor del parámetro ρ lo obtuvimos por validación cruzada dentro del grupo de entrenamiento. Para predecir el valor del óxido en el grupo de testeo utilizamos el modelo reemplazando la función de peso y el intercept por sus estimaciones. Esto nos permite obtener una estimación de su error cuadrático medio de predicción. Para tener una idea de su desempeño ajustamos un modelo de regresión lineal multivariada con estimaciones Ridge y Lasso

$$Y = \alpha + \sum_{j=1}^{301} \beta_j x(t_j). \tag{2}$$

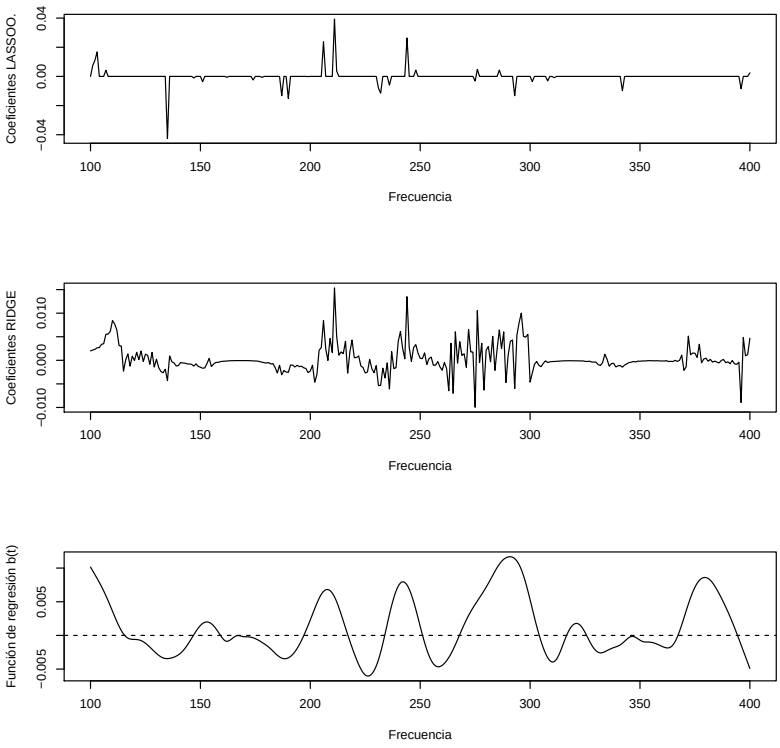

Figura 2: Estimaciones por Lasso, Ridge y MLF.

Este problema es análogo ya que si una grilla es regular y lo suficientemente fina uno espera que

$$\sum_{j=1}^{p} \beta_j x(t_j) \approx |I| \int_I b(t)x(t)\,dt,$$

donde $|I|$ es la longitud del dominio de las funciones.

Para los tres ajustes utilizamos el mismo grupo de entrenamiento y testeo. La función estimada y los valores de los coeficientes obtenidos por las metodologías pueden verse en la Figura 2. La Figura 3 muestra un boxplot que permite comparar los errores al predecir el óxido con cada metodología de estimación.

Los errores cuadráticos medios de predicción los presenta la siguiente tabla:

	Funcional	Lasso	Ridge
ECMP	0.65	1.01	1.16

Puede observarse que el estimador funcional supera ampliamente a los métodos multivariados. Si bien tanto Ridge como Lasso están pensados para situaciones de alta dimensionalidad ($p > n$) estos asumen poca correlación entre las covariables. En el caso de datos que vienen de una curva suave esta suposición no es razonable, uno espera una correlación fuerte entre valores $x(t_j)$ y $x(t_{j+k})$ si

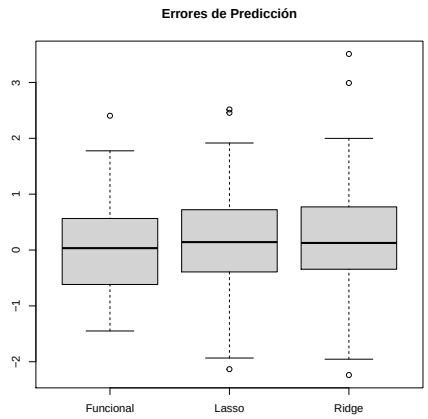

Figura 3: Boxplot de los errores por métodología de estimación.

la grilla es muy fina. Este podría dar una justificación de por qué el modelo funcional supera a las metodologías multivariadas en este conjunto de datos.

1.3. Estimación bajo restricciones

En esta tesis estudiamos el modelo lineal cuando existe información sobre la forma de la función de regresión. Cuando hablamos de forma, podemos imaginar hipótesis sobre su monotonía o concavidad, aunque en general nos referimos a que la función de peso pertenece a un subconjunto convexo del espacio de Hilbert donde se desarrolla el modelo. En la primera parte de este trabajo nos ocupamos en la estimación bajo restricciones. Es decir, estimación con funciones dentro de la clausura de dicho convexo. Para esto estudiamos una familia de estimadores de Sieves penalizados restringidos y probamos propiedades sobre muestras de tamaño fijo. A partir de estas obtenemos tasas de convergencias para el error de predicción y resultados de consistencia para la estimación de la función de regresión bajo la norma L_2. Los resultados que obtenemos son para conjuntos convexos en general y mostramos que las tasas son igual de buenas o mejores que las que se puedan obtener sin asumir restricciones. Implementamos los estimadores restringidos para restricciones de monotonía, concavidad y positividad. Realizamos una extensa simulación que muestra que los estimadores propuestos mejoran a los estimadores usuales para muestras de tamaño fijo.

En una segunda parte de esta tesis, consideramos el problema de estimar un subconjunto al cuál pertenece la función de regresión del modelo. Estudiamos el Modelo Lineal Histórico o Modelo Lineal Truncado donde la función de peso es constantemente nula a partir de un punto desconocido. Estudiamos la estimación del punto de cambio, es decir, el punto a partir del cual la función es cero. Obtenemos una tasa de convergencia para el estimador cuando en la literatura solo encontramos resultados sobre su consistencia. Mostramos además como obtener una estimación de la función de peso que se trunque en el punto de cambio estimado y obtenemos sus tasas pudiendo incluir restricciones de monotonía en dicha estimación. Realizamos una simulación computacional para estudiar el comportamiento de las metodologías para muestras finitas. Parte de estos resultados los obtuvimos en colaboración con M. Sued. Finalmente ilustramos el modelo y los estimadores sobre

un conjunto de datos de emisión de partículas sólidas.

Para abordar estos problemas estudiamos con detalle el modelo lineal sin restricciones y la convergencia de elementos aleatorios sobre espacios de Hilbert ampliando los resultados de la literatura para que se adecúen a nuestras necesidades. Algunos de los resultados obtenidos con este fin resultan de interés en sí mismos y dan herramientas teóricas para trabajar con datos funcionales más allá de los propósitos para los que se desarrollaron en esta tesis.

En la medida de lo posible presentamos los resultados en su forma más general para espacios de Hilbert separables.

1.4. Procesos estocásticos

En este apartado presentamos definiciones y resultados necesarios para el desarrollo de nuestro trabajo. Incluimos resultados sobre procesos estocásticos y análisis funcional siguiendo libros de referencia de cada tema. Como la mayoría de estos son considerados estandar los enunciamos sin demostración para completitud del manuscrito y del entorno de trabajo en el que se desarrolla la tesis.

Una de las primeras consideraciones que se debe tener al trabajar con datos funcionales es el espacio donde se desarrolla el modelo. Las opciones empiezan desde espacios métricos donde podemos medir distancias entre elementos, siguiendo por espacios de Banach donde se cuenta con una norma y las sucesiones de Cauchy son convergentes hasta espacios de Hilbert donde la norma proviene de un producto interno. Entre los espacios de Banach las opciones más frecuente son $\mathcal{C}[a,d]$ el espacio de las funciones continuas equipado con la norma del supremo o $L_p([a,d]) = \{f : [a,d] \to \mathbb{R},\ \int_a^d |f(t)|^p\, dt < +\infty\}$. Ambos espacios son separables y en el caso particular de $L_2([a,d])$ su norma viene inducida por el producto interno $\langle f, g\rangle = \int_I f(t)g(t)\, dt$. Otros ejemplos de espacios de Hilbert separables son $L_2[\mathbb{R}]$ o $H^k_{[a,d]} \subseteq L_2[a,d]$ el cual se define a partir del producto interno $\langle f, g\rangle_{H^k} = \sum_{i=0}^k \langle D^i f\ ,\ D^i g\rangle$ con D^i el operador derivada débil de orden i.

Los resultados de la primera parte de la tesis los obtenemos para cualquier espacio de Hilbert separable. Esto es diferente a buena parte de la literatura que suele trabajar con $L_2[a,d]$ o $\mathcal{C}[a,d]$

Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad, con Ω el espacio muestral, $\mathcal{A}$ una σ-álgebra y P una medida de probabilidad sobre $\mathcal{A}$ y un espacio medible $(E, \mathcal{B})$.

Definición 1.1 *(Proceso estocástico)*

Decimos que $X = (X_t,\, t \in T)$ es un proceso estocástico o proceso aleatorio, si X es una familia de variables aleatorias definidas sobre un espacio muestral $(\Omega, \mathcal{A}, P)$ que toman valores sobre el espacio de estados $(E, \mathcal{B})$ para un conjunto de indices T. Si $T \subseteq \mathbb{R}$ es un conjunto conexo, decimos que es un proceso estocástico (aleatorio) continuo. En el caso que $E = \mathbb{R}$ y $\mathcal{B} = \mathcal{B}_{\mathbb{R}}$ diremos que es un proceso aleatorio real.

Definición 1.2 *(Ley de X)*

Consideremos E^T el espacio de funciones de T a E y $\mathcal{S} = \sigma(\pi_t, t \in T)$ la sigma álgebra generada por las funciones $\pi_t : E^T \to E$ definidas como $\pi_t(x) = x(t)$, $x \in E^T$. La ley de X se define como

$$P_X(S) = P(X \in S),\ S \in \mathcal{S}.$$

Definición 1.3 *(Distribuciones finito dimensionales)*

Las distribuciones finito dimensionales de X son las distribuciones de los vectores aleatorios $(X_{t_1}, \ldots, X_{t_k})$ para cualquier $k \in \mathbb{N}$ y $t_1, \ldots, t_k \in T$.

Teorema 1.4 *Sea E un espacio métrico separable completo y $\mathcal{B}$ la sigma álgebra de Borel, las distribuciones finitos dimensionales de X determinan P_X.*

Definición 1.5 *Un proceso aleatorio real $X = (X_t, t \in T)$ se dice de segundo orden si $\mathtt{E}(X_t^2) < \infty$ para todo $t \in T$. En estos casos se puede definir la media del proceso*

$$m(t) = \mathtt{E}(X_t), \quad t \in T, \tag{3}$$

y la función de covarianza

$$c(s,t) = \mathtt{Cov}(X_s, X_t), \quad s, t \in T. \tag{4}$$

La función de covarianza satisface que es simétrica $c(s,t) = c(t,s)$ y positiva en el sentido que

$$\sum_{1 \leq i,j \leq k} \alpha_i \alpha_j c(t_i, t_j) \geq 0; \quad k \geq 1; \quad \alpha_1, \ldots, \alpha_k \in \mathbb{R}; \quad t_1, \ldots, t_k \in T.$$

En términos matriciales, la matriz $C = C_{t_1,\ldots,t_k} \in \mathbb{R}^{k \times k}$ definida por $(C)_{ij} = c(t_i, t_j)$ es simétrica y semidefinida positiva.

Definición 1.6 *Un proceso real de segundo orden $X = (X_t, t \in T)$ se dice Gaussiano si para cualquier $k \in \mathbb{N}$ y para todo $t_1, \ldots, t_k \in T$ el vector aleatorio $(X_{t_1}, \ldots, X_{t_k})$ tiene distribución normal multivariada con matriz de covarianza $C_{t_1,\ldots,t_k} \in \mathbb{R}^{k \times k}$.*

Es importante notar que por el Teorema 1.4 los procesos Gaussianos son determinados por m y c. A continuación damos dos ejemplos de procesos aleatorios que nos serán de utilidad más adelante.

Definición 1.7 *Un proceso de Wiener $W = (W_t, t \geq 0)$ es un proceso Gaussiano con $m(t) = 0$ y $c(s,t) = \sigma^2 \min(s,t)$. Donde σ^2 es una constante positiva.*

Definición 1.8 *Un proceso Ornstein-Uhlenbeck (O.U) es un proceso Gaussiano con función de media nula y función de covarianza $c(s,t) = \frac{\sigma^2}{\theta} e^{-\theta|s-t|}$ con σ^2, $\theta > 0$.*

Los procesos Ornstein-Uhlenbeck tienen trayectorias derivables mientras que los procesos de Wiener tienen trayectorias continuas pero no derivables.

Es bien conocido que las matrices de covarianza para vectores aleatorios son simétricas y semidefinidas positivas. El siguiente resultado que puede encontrarase en Bosq (2000) extiende este resultado a operadores de covarianza.

Teorema 1.9 *(Mercer) Si $c : [a,d] \times [a,d] \to \mathbb{R}$ es una función de covarianza continua, entonces existe una sucesión (e_n) completa de funciones continuas que forman una base ortonormal de $L_2[a,d]$ y además, una secuencia decreciente (λ_n) de reales no negativos tales que*

$$c(s,t) = \sum_{n=1}^{\infty} \lambda_n e_n(s) e_n(t), \quad s, t \in [a,d]. \tag{5}$$

La convergencia de esta serie es uniforme sobre $[a,d]^2$ *y se satisface*

$$\int_a^d c(t,t)\,dt = \sum_{n=1}^{\infty} \lambda_n < \infty. \tag{6}$$

Más aún,

$$\int_a^d c(s,t)e(s)\,ds = \lambda_n e_n(t), \quad t \in [a,d], \quad n \in \mathbb{N}. \tag{7}$$

Si definimos el operador lineal $T : C[a,d] \to C[a,d]$, $T(f) = \int_a^d c(s,t)f(s)\,ds$ los pares e_n, λ_n resultan sus autofunciones y autovalores asociados. El siguiente Teorema extiende las nociones de componentes principales a procesos estocásticos.

Teorema 1.10 *(Expansión de Karhunen-Loeve (K.L.)) Sea* $X = (X_t,\ t \in [a,d])$ *un proceso de segundo orden con función de covarianza continua. Tenemos*

$$X_t = m(t) + \sum_{n=1}^{\infty} \chi_n e_n(t), \quad t \in [a,d], \tag{8}$$

con (χ_n) *una secuencia de variables aleatorias reales con esperanza cero,* $\mathtt{E}(\chi_n\chi_k) = \lambda_n\delta_{n,k},\ n,k \in \mathbb{N}$ *y los pares* e_n, λ_n *los definidos por el Teorema 1.9.*

Dem: Ver Teorema 1.5 de Bosq (2000).

1.5. Procesos aleatorios sobre un espacio de Hilbert

Para definir elementos aleatorios definidos sobre un espacio de Hilbert separable hacemos un repaso de algunas definiciones y resultados que necesitamos del análisis funcional. Empezamos por clasificar los operadores lineales y comentar algunas de sus propiedades. Finalmente enunciamos un Teorema de descomposición espectral.

En lo que sigue $\mathbb{H}$ es un espacio de Hilbert separable con norma $\|u\| = \sqrt{\langle u,u\rangle}$.

1.5.1. Clasificación de operadores lineales

Definición 1.11 $T : \mathbb{H} \to \mathbb{H}$ *es un operador acotado si existe* $C > 0$ *tal que*

$$\sup_{\|u\|\le 1} \|T(u)\| \le C.$$

Los operadores lineales acotados son continuos. El espacio de operadores acotados es un espacio de Banach con la norma del supremo $\|T\| = \sup_{\|u\|\le 1} \|T(u)\|$.

Definición 1.12 *Un operador* T *se dice compacto si la imagen sobre cualquier conjunto acotado tiene clausura compacta.*

El espacio de operadores lineales compactos es un espacio de Banach con la norma del supremo y es un subconjunto del espacio de operadores acotados.

Proposición 1.13 *Un operador T es compacto si y solo se puede escribir de la forma*

$$T(u) = \sum_{n=1}^{\infty} \lambda_n \langle u, f_n \rangle e_n. \tag{9}$$

con $\{f_n\}$ y $\{e_n\}$ conjuntos ortonormales de $\mathbb{H}$ y $\{\lambda_n\}$ una secuencia de números reales no negativos tales con $\lambda_n \to 0$.

Definición 1.14 *Un operador lineal compacto se dice de Hilbert-Schmidt si en la descomposición dada por la Ecuación 9 satisface*

$$\sum_{n=1}^{\infty} \lambda_j^2 < +\infty.$$

Definición 1.15 *Un operador lineal compacto se dice Nuclear si en la descomposición dada por la Ecuación 9 satisface*

$$\sum_{n=1}^{\infty} |\lambda_j| < +\infty.$$

Definición 1.16 *El operador $T : \mathbb{H} \to \mathbb{H}$ se dice:*

Simétrico o autoadjunto si $\langle u, Tv \rangle = \langle Tu, v \rangle$ para todo $u, v \in \mathbb{H}$.

Positivo si $\langle T(u), u \rangle \geq 0$ para todo $u \in \mathbb{H}$.

Teorema 1.17 *(Diagonalización de un operador compacto, simétrico y positivo)*

Si $T : \mathbb{H} \to \mathbb{H}$ es un operador lineal compacto, simétrico y positivo, existe una secuencia de números reales no negativos λ_n tal que $\lambda_n \to 0$ cuando $n \to \infty$, y una base ortonormal $\{e_n\}$ para $Nu(T)^{\perp}$ tal que para todo $u \in \mathbb{H}$

$$T(u) = \sum_{n=1}^{\infty} \lambda_n \langle u, e_n \rangle e_n. \tag{10}$$

Si $Nu(T)^{\perp}$ es un espacio de dimensión finita la serie es una suma finita. Como $T(e_n) = \lambda_n e_n$ se dice que $\{\lambda_n\}$ y $\{e_n\}$ son los autovalores y autovectores de T respectivamente.

La demostración se desprende como corolario de la diagonalización de operadores compactos autoadjuntos. Ver Teorema 5.1, Capítulo 4, Conway (1990).

Corolario 1.18 *Bajo las hipótesis del Teorema 1.17, existe $T^{1/2} : \mathbb{H} \to \mathbb{H}$, simétrico, positivo y compacto tal que $T = T^{1/2} \circ T^{1/2}$ y se define como*

$$T^{1/2}(u) = \sum_{n=1}^{\infty} \lambda_n^{1/2} \langle u, e_n \rangle e_n. \tag{11}$$

Además, si T es un operador Nuclear, $T^{1/2}$ resulta un operador de Hilbert Schmidt.

Teorema 1.19 *(min-max) Sea $H \subseteq \mathbb{H}$ un subespacio de dimensión finita p. Sea $D : H \to \mathbb{H}$ y $0 \leq \tau_1 \leq \ldots \leq \tau_p$ los autovalores de $D^\star D$. Resulta*

$$\tau_k = \min_{\substack{U \subseteq H \\ dim(U)=k}} \max_{u \in U} \left\{ \|Du\|^2 \quad con \quad \|u\| \leq 1 \right\}. \tag{12}$$

1.5.2. Elementos aleatorios en un espacio de Hilbert separable

Definición 1.20 *Decimos que $x : (\Omega, \mathcal{A}) \to (\mathbb{H}, \mathcal{B}_{\mathbb{H}})$ es un $\mathbb{H}$ elemento aleatorio o una $\mathbb{H}$ variable aleatoria si $\langle u, x \rangle$ es una variable aleatoria real para todo $u \in \mathbb{H}$.*

Definición 1.21 *Sea x un $\mathbb{H}$ elemento aleatorio que satisface $\mathtt{E}(\|x\|) < +\infty$. Definimos la esperanza de x y notamos $\mathtt{E}(x)$ al elemento de $\mathbb{H}$ que satisface*

$$\mathtt{E}(\langle u, x \rangle) = \langle u, \mathtt{E}(x) \rangle, \quad u \in \mathbb{H}. \tag{13}$$

La buena definición se desprende del Teorema de Representación de Riesz ya que $u \to \mathtt{E}(\langle u, x \rangle)$ es una transformación lineal acotada de $\mathbb{H}$ en $\mathbb{R}$. La Esperanza de $\mathbb{H}$ elementos aleatorios mantiene muchas de las propiedades usuales conocidas para variables aleatorias reales. Por ejemplo

$$\|\mathtt{E}(x)\| \leq \mathtt{E}\left(\|x\|\right).$$

Además, si $\mathbb{H}_1$ y $\mathbb{H}_2$ son dos espacios de Hilbert separables y T es un operador lineal acotado de $\mathbb{H}_1$ a $\mathbb{H}_2$ entonces

$$\mathtt{E}(T(x)) = T(\mathtt{E}(x)).$$

Definición 1.22 *(Operador de covarianza)*

Sea x un $\mathbb{H}$ elemento aleatorio con $\mathtt{E}(x) = m$ y $\mathtt{E}(\|x\|^2) < \infty$. Definimos el operador de covarianza $\Gamma : \mathbb{H} \to \mathbb{H}$

$$\Gamma(u) = \mathtt{E}\left(\langle x - m\,,\, u \rangle (x - m)\right).$$

Teorema 1.23 *Un operador $\Gamma : \mathbb{H} \to \mathbb{H}$ es de covarianza si y sólo si es simétrico positivo y nuclear.*

Es decir, Γ es un operador de covarianza para algún $\mathbb{H}$ elemento aleatorio si y sólo si, existe una base ortonormal de $Nu(\Gamma)^\perp$, $\{e_n\}$ y una sucesión decreciente de números no negativos $(\lambda_n)_{n \in \mathbb{N}}$ tales que $\sum_{n=1}^{\infty} \lambda_n < \infty$ y

$$\Gamma(u) = \sum_{n=1}^{\infty} \lambda_n \langle u, e_n \rangle e_n. \tag{14}$$

Si la dimensión de $Nu(\Gamma)^\perp$ es finita también lo es la sumatoria.

Este teorema puede encontrase en el Capítulo 1 de Bosq (2000) (Teorema 1.7) y es análogo al Teorema de Mercer 1.9 sobre un espacio de Hilbert. Al no pensar los elementos del Hilbert como funciones se pierde la noción de convergencia uniforme y de continuidad en elementos de la base.

El siguiente resultado es la expansión de Karhunen Loeve enunciada sobre un espacio de Hilbert separable. El resultado es análogo al Teorema 1.17. Como no lo encontramos enunciado Bosq (2000) ofrecemos una demostración análoga a la encontrada en la referencia.

Teorema 1.24 *(Representación de un $\mathbb{H}$ elemento aleatorio)*

Sea x un $\mathbb{H}$ elemento aleatorio tal que $\mathtt{E}(\|x\|^2) < +\infty$, con esperanza m y operador de covarianza Γ. Existe $(\chi_n)_{n\in\mathbb{N}}$ una secuencia de variables aleatorias reales con media cero que satisface $\mathtt{E}(\chi_n\chi_k) = \lambda_n\delta_{n,k}$ tal que

$$x = m + \sum_{n=1}^{\infty} \chi_n e_n, \tag{15}$$

con λ_n y e_n los autovalores y autofunciones definidos en el Teorema 1.23.

Más aún, $m = \mathtt{E}(x)$, $\chi_n = \langle x - m\,,\, e_n\rangle$ y

$$\sum_{n=1}^{\infty} \lambda_n = \mathtt{E}\left(\|x - m\|^2\right). \tag{16}$$

Dem: 7.1

1.5.3. El proceso aleatorio como dato funcional

Cuando el proceso aleatorio es un elemento del espacio $L_2(I) = \{f : I \subseteq \mathbb{R}^d \to \mathbb{R},\ \int_I f(t)^2\,dt < \infty\}$ podemos interpretarlo como una curva aleatoria. A continuación, mostramos que las definiciones de media y operador de covarianza coinciden con las de procesos aleatorios.

Lema 1.25 *Sea $I \subseteq \mathbb{R}$ un intervalo y x un $L_2(I)$ elemento aleatorio tal que $\mathtt{E}(\|x\|) < +\infty$. Luego la esperanza del proceso existe y satisface*

$$\mathtt{E}(x)(t) = \mathtt{E}(x(t)), \quad t \in I. \tag{17}$$

Dem 7.2.

Lema 1.26 *Sea x un $L_2([a,d])$ elemento aleatorio tal que $\mathtt{E}\left(\|x\|_\infty^2\right) < +\infty$ entonces*

$$\Gamma[u](s) = \int_a^d \mathtt{Cov}\left(x(s), x(t)\right) u(t)\,dt. \tag{18}$$

Más aún, si $(e_n)_{n\in\mathbb{N}}$ y $(\lambda_n)_{n\in\mathbb{N}}$ son los definidos por el Teorema 1.9 tal que

$$\mathtt{Cov}(x(s), x(t)) = \sum_{n=1}^{\infty} \lambda_n e_n(s) e_n(t).$$

Entonces $(e_n)_{n\in\mathbb{N}}$ y $(\lambda_n)_{n\in\mathbb{N}}$ son los definidos en el Teorema 1.24 para Γ.

Dem 7.3.

2. Convergencia de elementos aleatorios en un espacio de Hilbert

En esta sección desarrollamos herramientas teóricas que necesitamos para determinar las propiedades asintóticas de los estimadores que trabajamos en esta tesis. Como resultado principal obtenemos una ley de los grandes números uniformes que nos permite estudiar la convergencia de sucesiones de $\mathbb{H}$ elementos aleatorios.

A continuación damos un breve resumen de los tipos de convergencia con los que trabajamos a lo largo de la tesis.

Definición 2.1 *(Convergencias)*

Decimos que u_n converge en norma a u y notamos $u_n \to u$ si

$$\|u - u_n\| \to 0.$$

Decimos que u_n converge débilmente a u y notamos $u_n \rightharpoonup u$ si para todo $v \in \mathbb{H}$

$$\langle u_n, v\rangle \to \langle u, v\rangle.$$

Si T es un operador lineal, decimos que u_n converge bajo T a u y notamos $u_n \xrightarrow{T} u$ si

$$T(u_n) \to T(u).$$

Es importante notar que no hay unicidad del límite para la convergencia bajo T si este no es inyectivo. En efecto, si $u_n \xrightarrow{T} u$ y $w \in Nu(T)$ entonces $u_n \xrightarrow{T} u + w$. Es un hecho bien conocido que la convergencia en norma implica la convergencia débil y que los operadores acotados preservan la convergencia en norma.

El siguiente resultado que se encuentra en el Capítulo 6 Conway (1990), Proposición 3.3, nos dice que convergencia débil implica convergencia bajo T.

Proposición 2.2 *Si T es un operador lineal compacto y $u_n \rightharpoonup u$ entonces $u_n \xrightarrow{T} u$.*

La siguiente proposición nos dice que si T es inyectivo, la convergencia bajo T implica convergencia débil para sucesiones acotadas en norma.

Proposición 2.3 *Si T es un operador lineal compacto tal que $u_n \xrightarrow{T} u$ y $\sup_{n\in\mathbb{N}} \|u_n\| < +\infty$ entonces para todo $v \in Nu(T)^{\perp}$ se tiene que $\langle u_n, v\rangle \to \langle u, v\rangle$.*

Dem: 8.1

Observación 2.4 *La condición $\sup_{n\in\mathbb{N}} \|u_n\| < \infty$ es necesaria ya que por el Teorema de Banach-Steinhaus, (Capítulo 3, Conway (1990)) las sucesiones débilmente convergentes están acotadas en norma. Si T es compacto e inyectivo, la sucesión $u_n = \lambda_n^{-1/2} e_n$, donde λ_n y e_n son los definidos en el Teorema 1.17, satisface $T(u_n) = \lambda_n^{1/2} e_n \to T(0)$. Como u_n no está acotado en norma no puede converger débilmente.*

2.1. Ley de los Grandes Números Uniforme

En lo que sigue x es un $\mathbb{H}$ elemento aleatorio con operador de covarianza Γ con descomposición dada como en el Teorema 1.23.

Definición 2.5 *Llamamos Γ-Norma, a la (semi) norma asociada al operador Γ*

$$\|u\|_\Gamma := \|\Gamma^{1/2}(u)\| = \left(\sum_{n=1}^{\infty} \lambda_n \langle u, e_n\rangle^2\right)^{1/2}. \tag{19}$$

Notemos que resulta una norma si y solo si Γ es inyectivo. La siguiente observación es consecuencia directa del Teorema 1.24.

Observación 2.6 *Si u es un elemento fijo en $\mathbb{H}$ la Γ-Norma cumple*

$$\|u\|_\Gamma^2 = \mathtt{E}(\langle u, x - m\rangle^2). \tag{20}$$

En el Capítulo 3 en la Proposición 3.6 mostramos la importancia de la Γ-Norma ya que aparece en la tasa de convergencia del error de predicción del modelo lineal.

Observación 2.7 *Si Γ es inyectivo el par $(\mathbb{H}, \|\cdot\|_\Gamma)$ forma un espacio normado con producto interno asociado $\langle u, v\rangle_\Gamma = \langle \Gamma^{1/2}(u), \Gamma^{1/2}(v)\rangle$. Este espacio no es completo.*

Dem: 8.2

Nos resulta de interés estimar la Γ-Norma de un elemento $u \in \mathbb{H}$ a partir de una muestra $x_1, \ldots, x_n$. La siguiente definición surge naturalmente al interpretar la (semi) norma mediante la Ecuación 20.

Definición 2.8 *(Γ-Norma empírica) Si notamos $\overline{x} = \frac{1}{n}\sum_{i=1}^n x$ la norma empírica se define como*

$$\|u\|_{\Gamma_n} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \langle u, x_i - \overline{x}\rangle^2}. \tag{21}$$

Si bien la Ley de los Grandes Números dice que para cada $u \in \mathbb{H}$ se tiene $\|u\|_{\Gamma_n}^2 \xrightarrow{P} \|u\|_\Gamma^2$, necesitamos un resultado uniforme. En particular vamos a estudiar el comportamiento de la siguiente variable aleatoria

$$\sup_{\|u\|\leq 1} \left| \|u\|_\Gamma^2 - \|u\|_{\Gamma_n}^2 \right|.$$

Para esto utilizamos las siguientes condiciones que pedimos como hipótesis sobre una muestra aleatoria $x_1, \ldots, x_n$ no necesariamente independiente.

Hipótesis sobre la muestra aleatoria.

Sea $\mathbb{H}$ un espacio de Hilbert separable y $x_1, \ldots, x_n$ una muestra aleatoria idénticamente distribuida sobre $\mathbb{H}$ con media m y operador de covarianza Γ.

(A1) Existe una constante $C < \infty$ tal que para todo r, $s \in \mathbb{N}$

$$\mathtt{Var}\left(\frac{1}{n}\sum_{i=1}^{n}\langle x_i - m, e_r\rangle\langle x_i - m, e_s\rangle\right) \leq \frac{C}{n}\mathtt{E}\left(\langle x - m, e_r\rangle^2\right)\mathtt{E}\left(\langle x - m, e_s\rangle^2\right).$$

(A2) Para una secuencia $\mathtt{m}_n$ de números reales no negativos que tiende a cero

$$\mathtt{E}\left(\|m - \overline{x}\|^2\right) = \mathcal{O}\left(\mathtt{m}_n\right).$$

(A3) Para una secuencia $\mathtt{l}_n$ de números reales no negativos que tiende a cero

$$\mathtt{E}\left(\|m - \overline{x}\|^4\right) = \mathcal{O}\left(\mathtt{l}_n^2\right).$$

La siguiente observación nos muestra que estas hipótesis son muy poco restrictivas y que son consecuencias inmediatas de las hipótesis habitualmente utilizadas en la literatura.

Observación 2.9 *La condición* $(A1)$ *se satisface cuando los* x_i *son elementos gaussianos independientes.*

Si $\mathtt{E}(\|x - m\|^4) < \infty$ *y la muestra es independiente, la condición* $(A2)$ *se obtiene con* $\mathtt{m}_n = \frac{1}{n}$ *y la condición (A3) vale tomando* $\mathtt{l}_n = \frac{1}{\sqrt{n}}$.

$$\mathtt{E}\left(\|\overline{x} - m\|^2\right) = \frac{1}{n^2}\sum_{i,j=1}^{n}\mathtt{E}\left(\langle x_i - m, x_j - m\rangle\right) = \frac{1}{n}\mathtt{E}\|x_1 - m\|^2 = \frac{1}{n}\sum_{j=1}^{\infty}\lambda_j,$$

$$\mathtt{Var}\left(\|\overline{x} - m\|^2\right) = \mathtt{Var}\left(\frac{1}{n^2}\|\sum_{i=1}^{n}(x_i - m)\|^2\right) \leq \mathtt{Var}\left(\frac{n}{n^2}\sum_{i=1}^{n}\|x_i - m\|^2\right) = \frac{1}{n}\mathtt{Var}\left(\|x - m\|^2\right).$$

Para más propiedades sobre la media de la muestra se puede consultar Grenander (1981). En particular, en el Teorema 7 del Capítulo 4, se establece un Teorema Central del Límite para $\overline{x}$.

El siguiente Teorema establece el comportamiento de la la norma empírica uniformemente sobre la bola cerrada de radio uno. El resultado es muy fuerte ya que la bola cerrada no es compacta en un espacio de Hilbert de dimensión infinita. Las ideas que usamos para probarlo se basan en parte de la demostración de las tasas del estimador de Crambes et al. (2009). Creemos que es un aporte valioso a la teoría de datos funcionales que no hemos encontrado expresado de manera compacta ni referenciable.

Teorema 2.10 *(Ley de los grandes números uniforme)*

Sea $x_1 \ldots x_n$ *una muestra de* $\mathbb{H}$ *elementos aleatorios que satisface las condiciones (A1) y (A2) con operador de covarianza* Γ *con descomposición como en el Teorema 1.23.*

Sean $\mathtt{s}_k = \sum_{j=k+1}^{\infty}\lambda_j$ *y* $\mathtt{t}_k = \sum_{l=1}^{k}\mathtt{s}_l$.

Sea k_n *una sucesión de números naturales tal que* $\lim_{n\to\infty}\dfrac{\mathtt{t}_{k_n}}{n} = 0$.

Definimos para cada $u \in \mathbb{H}$ *la sucesión* $\mathtt{f}_n(u) = \max\left\{\sqrt{\frac{\mathtt{t}_{k_n}}{n}}\|u\|_\Gamma\,,\ \frac{\mathtt{s}_{k_n}}{\sqrt{n}}, \mathtt{m}_n\right\}$, *entonces*

$$\mathtt{E}\left(\sup_{\|u\|\leq 1}\frac{1}{\mathtt{f}_n(u)}\left|\|u\|_\Gamma^2 - \|u\|_{\Gamma_n}^2\right|\right) = \mathcal{O}\left(1\right). \tag{22}$$

Dem: 8.3

El siguiente Corolario establece el mismo resultado pero en probabilidad.

Corolario 2.11 *(Ley débil de los grandes números uniforme)*

Con las mismas definiciones e hipótesis del Teorema 2.10 salvo la condición (A2) sustituida por la condición mas leve

$$\|m - \overline{x}\|^2 = \mathcal{O}_P(\mathtt{m}_n).$$

Se obtiene

$$\sup_{\|u\|\leq 1}\frac{1}{\mathtt{f}_n(u)}\left|\|u\|_\Gamma^2 - \|u\|_{\Gamma_n}^2\right| = \mathcal{O}_P\left(1\right). \tag{23}$$

La demostración de este resultado es análoga a la del Teorema 2.10.

El siguiente resultado permite controlar de manera uniforme el segundo momento del supremo de la norma empírica sobre la bola cerrada de radio uno.

Corolario 2.12 *Con las mismas definiciones e hipótesis del Teorema 2.10, si además vale la condición (A3) y definimos* $\mathtt{g}_n(u) = \max\left\{\sqrt{\frac{\mathtt{t}_{k_n}}{n}}\|u\|_\Gamma\,,\ \frac{\mathtt{s}_{k_n}}{\sqrt{n}}, \mathtt{l}_n\right\}$ *tenemos que*

$$\mathtt{E}\left(\sup_{\|u\|\leq 1}\frac{1}{\mathtt{g}_n^2(u)}\left|\|u\|_\Gamma^2 - \|u\|_{\Gamma_n}^2\right|^2\right) = \mathcal{O}\left(1\right). \tag{24}$$

En particular, si la muestra es independiente

$$\mathtt{E}\left(\sup_{\|u\|\leq 1}\left|\|u\|_\Gamma^2 - \|u\|_{\Gamma_n}^2\right|^2\right) = \mathcal{O}\left(\frac{1}{n}\right).$$

Dem: 8.4

Observación 2.13 *En el Teorema 2.10 siempre es posible tomar* $k_n = \mathcal{O}(n)$ *ya que* $\lim_{n\to+\infty}\dfrac{\mathtt{t}_n}{n} = 0$. *Dem: 8.5.*

El siguiente teorema nos permite obtener tasas en Γ-Norma a partir de las tasas en norma empírica y es nuestra principal herramienta al estudiar el comportamiento asintótico de los estimadores propuestos en esta tesis.

Teorema 2.14 *Con las mismas definiciones e hipótesis del Teorema 2.10 sea u_n una secuencia de $\mathbb{H}$ elementos aleatorios entonces*

$$|\|u_n\|_\Gamma^2 - \|u_n\|_{\Gamma_n}^2| = \mathcal{O}_P\left(\text{máx}\left\{\frac{\mathtt{t}_{k_n}}{n}\|u_n\|^2,\, \frac{\mathtt{s}_{k_n}}{\sqrt{n}}\|u_n\|^2,\, \mathtt{m}_n\|u_n\|^2\right\}\right). \tag{25}$$

Si $\|u_n\|_{\Gamma_n}^2 = \mathcal{O}_P(\mathtt{a}_n)$ entonces

$$\|u_n\|_\Gamma^2 = \mathcal{O}_P\left(\text{máx}\left\{\mathtt{a}_n,\, \frac{\mathtt{t}_{k_n}}{n}\|u_n\|^2,\, \frac{\mathtt{s}_{k_n}}{\sqrt{n}}\|u_n\|^2,\, \mathtt{m}_n\|u_n\|^2\right\}\right) \tag{26}$$

Dem: 8.6

La siguiente proposición juega un rol parecido al anterior pero para las tasas en esperanza.

Proposición 2.15 *Con las mismas definiciones e hipótesis del Teorema 2.10 sea u_n una secuencia de $\mathbb{H}$ elementos aleatorios tales que $E(\|u_n\|_{\Gamma_n}^2) = \mathcal{O}(\mathtt{a}_n)$*

Si se satisface $\sup \mathtt{E}(\|u_n\|^2) < +\infty$*, entonces*

$$\mathtt{E}(\|u_n\|_\Gamma) = \mathcal{O}(\text{máx}\{\mathtt{a}_n^{1/2},\, n^{-1/4},\, \mathtt{m}_n^{1/2}\}). \tag{27}$$

Si también vale $\sup \mathtt{E}(\|u_n\|^4) < +\infty$ *y se cumple la condición (A3) entonces*

$$\mathtt{E}(\|u_n\|_\Gamma^2) = \mathcal{O}(\text{máx}\{\mathtt{a}_n,\, n^{-1/2},\, \mathtt{l}_n\}). \tag{28}$$

Dem: 8.7

Concluimos esta serie de resultados observando que las tasas de convergencia de una sucesión de $\mathbb{H}$ elementos aleatorios se pueden obtener de las tasas en norma empírica siempre que la sucesión de sus normas esté acotada en probabilidad.

Observación 2.16 *Con las definiciones e hipótesis del Teorema 2.14 si $\|u_n\|^2 = \mathcal{O}_P(1)$ entonces para toda sucesión k_n tal que $\mathtt{t}_{k_n} = o(n)$*

$$\|u_n\|_\Gamma^2 = \mathcal{O}_P\left(\text{máx}\left\{\mathtt{a}_n,\, \frac{\mathtt{t}_{k_n}}{n},\, \frac{\mathtt{s}_{k_n}}{\sqrt{n}},\, \mathtt{m}_n\right\}\right).$$

Como siempre existe k_n tal que $\mathtt{s}_{k_n} = \mathcal{O}(n^{-1/2})$, si $sup_{n\in\mathbb{N}}\mathtt{t}_n < \infty$ entonces

$$\|u_n\|_\Gamma^2 = \mathcal{O}_P\left(\text{máx}\left\{\mathtt{a}_n,\, \frac{1}{n},\, \mathtt{m}_n\right\}\right).$$

Si en cambio $sup_{n\in\mathbb{N}}\mathtt{t}_n = \infty$ la tasa tendrá un término de orden mayor a n^{-1}.

Finalmente, si $\|u_n\|_{\Gamma_n} \xrightarrow{P} 0$, $\|u_n\|^2 = \mathcal{O}_P(\sqrt{n})$ y $\mathtt{m}_n = o(\sqrt{n})$ entonces

$$\|u_n\|_\Gamma \xrightarrow{P} 0.$$

A continuación mostramos dos ejemplos donde según la elección de distintos k_n puede variar el resultado de las tasas que se obtienen por el Teorema 2.14.

Ejemplo 2.17

1) Si $\lambda_k = \mathcal{O}(k^{-1-q})$ con $q > 0$ entonces $\mathtt{s}_k = k^{-q}$ y obtenemos

$$\frac{s_{k_n}}{\sqrt{n}} = \frac{{k_n}^{-q}}{\sqrt{n}}, \qquad \frac{\mathtt{t}_{k_n}}{n} = \begin{cases} \mathcal{O}\left(\frac{k_n^{-q+1}}{n}\right) & si\ q < 1, \\ \mathcal{O}\left(\frac{\ln(k_n)}{n}\right) & si\ q = 1, \\ \mathcal{O}\left(\frac{1}{n}\right) & si\ q > 1. \end{cases} .$$

La siguiente tabla muestra los valores de $\text{máx}\left\{\frac{s_{k_n}}{\sqrt{n}}, \frac{t_{n_k}}{n}\right\}$ *para distintos valores de q y k_n*

	$0 < q < 1$	$q = 1$	$q > 1$
$k_n = cte$	$n^{-1/2}$	$n^{-1/2}$	$n^{-1/2}$
$k_n = n$	n^{-q}	$n^{-1}\ln(n)$	n^{-1}
$k_n = n^{1/2q}$	$n^{-3/2+1/2q}$	$n^{-1}\ln(n)$	n^{-1}

Es interesante notar que si $q \leq 1/3$ la sucesión $k_n = n^{1/2q}$ tiende a infinito demasiado rápido causando que $\text{máx}\{\frac{s_{k_n}}{\sqrt{n}}, \frac{t_{n_k}}{n}\} \geq 1$.

2) En el caso mas extremo con $\lambda_k = \mathcal{O}\left(\frac{1}{k\ln^2(k)}\right)$, resulta

$$\frac{\mathtt{s}_{k_n}}{\sqrt{n}} = \mathcal{O}\left(\frac{1}{\sqrt{n}\ln(k_n)}\right) \qquad y \qquad \frac{\mathtt{t}_{k_n}}{n} = \mathcal{O}\left(\frac{k_n}{n\ln(k_n)}\right).$$

	$\text{máx}\left\{\frac{s_{k_n}}{\sqrt{n}}, \frac{t_{n_k}}{n}\right\}$
$k_n = cte$	$n^{-1/2}$
$k_n = n^{1/2}$	$n^{-1/2}\ln^{-1}(n)$
$k_n = n$	$\ln^{-1}(n)$

Dem: 8.8

2.2. Relaciones entre distintos tipos de convergencia.

En esta subsección mostramos algunas implicaciones entre los distintos tipos de convergencias. En las Figuras 4 y 5 presentamos diagramas que resumen los resultados obtenidos.

La siguiente proposición nos da consistencia para la convergencia débil a partir de la consistencia en Γ-Norma.

Proposición 2.18 *Sea Γ un operador de covarianza y u_n una secuencia de $\mathbb{H}$ elementos aleatorios que cumplen que $\|u_n\|_\Gamma \xrightarrow{P} 0$ y $\sup_{n\in\mathbb{N}} \mathtt{E}\left(\|u_n\|^{2r}\right) < +\infty$, para algún $r \in \mathbb{N}$. Resulta*

$$\lim_{n\to+\infty} \mathtt{E}\left(\|u_n\|_\Gamma^r\right) = 0. \tag{29}$$

Más aún, para todo $v \in Nu^\perp(\Gamma)$ vale que

$$\lim_{n\to+\infty} \mathtt{E}(\langle u_n, v\rangle) = 0.$$

Dem 8.9

El siguiente Corolario nos da condiciones bajo las cuales la convergencia en probabilidad de la Γ-Norma implica la convergencia de la esperanza de la norma cuadrada.

Proposición 2.19 *Sea u_n es una secuencia de $\mathbb{H}$ elementos aleatorios tales que para un $u_0 \in \mathbb{H}$ se verifica $\|u_n - u_0\|_\Gamma \xrightarrow{P} 0$. Si además $\limsup_n \mathtt{E}(\|u_n\|^2) \le \|u_0\|^2$, entonces*

$$\lim_{n\to+\infty} \mathtt{E}\left(\|u_n - u_0\|_\Gamma^2\right) = \lim_{n\to+\infty} \mathtt{E}(\|u_n - u_0\|^2) = 0.$$

Dem: 8.10

Continuamos con un resultado que muestra como las tasas de la convergencia en Γ-Norma condicionan las tasas de la convergencia débil sobre un conjunto amplio de funciones. Estas tasas no son necesariamente óptimas.

Sea $\mathtt{v}_k$ una sucesión de reales positivos, llamamos

$$B(\mathtt{v}, C) = \{vu \in Nu(\Gamma)^\perp,\ \|u\| \le 1 \ :\ |\langle v\,,\, e_k\rangle| \le C\mathtt{v}_k\}. \tag{30}$$

Proposición 2.20 *Sea u_n una secuencia de $\mathbb{H}$ elementos aleatorios tales que $\|u_n\|_\Gamma = \mathcal{O}_P(\mathtt{b}_n)$ y $\|u_n\| = \mathcal{O}_P(1)$. Sean e_k las autofunciones asociadas a Γ. Si llamamos $\mathtt{w}_k = \sqrt{\sum_{j=k+1}^\infty \mathtt{v}_k^2}$, para cualquier sucesión k_n que tiende a infinito*

$$\sup_{u\in\mathcal{B}(\mathtt{v},C)} |\langle u_n\,,\, u\rangle| = \mathcal{O}_P\left(\mathtt{b}_n\lambda_{k_n}^{-1/2} + \mathtt{w}_{k_n}\right).$$

Si $\mathtt{E}(\|u_n\|_\Gamma) = \mathcal{O}(\mathtt{b}_n)$ entonces

$$\mathtt{E}\left(\sup_{u\in\mathcal{B}(\mathtt{v},C)} |\,\langle u_n\,,\, u\rangle\,|\right) = \mathcal{O}\left(\mathtt{b}_n\lambda_{k_n}^{-1/2} + \mathtt{E}(\|u_n\|)\mathtt{w}_{k_n}\right).$$

Dem: 8.11

El siguiente Corolario nos dice como conseguir tasas para la convergencia en norma a partir de las tasas en Γ-Norma y los autovalores del operador de covarianza de la covariable.

Corolario 2.21 *Sea u_n una secuencia de $\mathbb{H}$ elementos aleatorios y $\mathtt{b}_n, \mathtt{d}_n$ dos sucesiones de números no negativos tales que $\|u_n - u_0\|_\Gamma = \mathcal{O}_P(\mathtt{b}_n)$ y $\mathtt{E}(\|u_n\|^2) \leq \|u_0\|^2 + \mathtt{d}_n$. Si llamando $\mathtt{w}_k = \sqrt{\sum_{r=k+1}^{\infty} \langle u_0 \, , \, e_r \rangle^2}$ tenemos para cualquier $k_n \in \mathbb{N}$ que*

$$\mathtt{E}(\|u_n - u_0\|^2) = \mathcal{O}\left(\mathtt{d}_n + \mathtt{b}_n \lambda_{k_n}^{-1/2} + \mathtt{w}_{k_n}^2\right).$$

Dem: 8.12

Las siguientes figuras presentan un diagrama de los resultados obtenidos. Cada flecha apunta al resultado que se puede obtener a partir de las condiciones de donde sale.

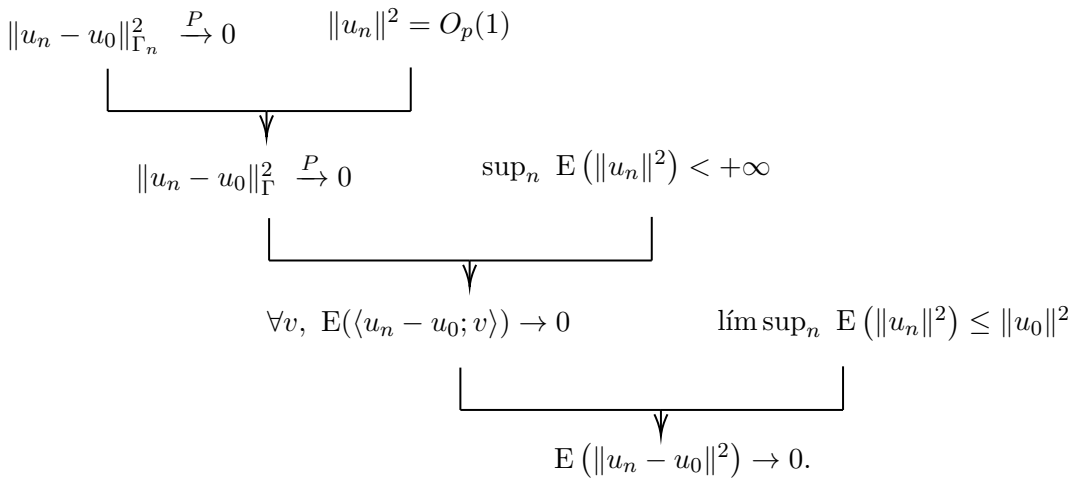

Figura 4: Diagrama de implicaciones de consistencias.

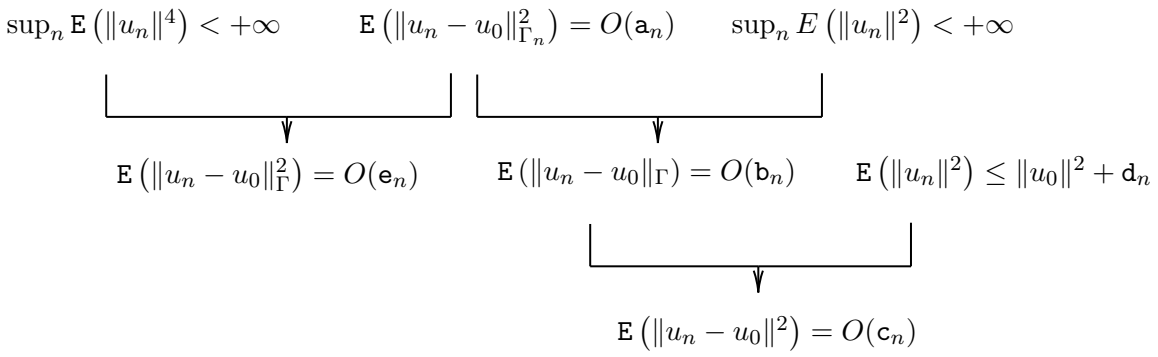

Figura 5: Diagrama de implicaciones de tasas.

3. Modelo lineal con covariable funcional y respuesta escalar

Uno de los primeros resultados sobre el modelo lineal funcional con respuesta escalar fue presentado en Cardot et al. (1999) donde se estudió la consistencia del estimador basado en componentes principales funcionales. Este estimador, que notamos $\widehat{b}_{pca}$, se obtiene como el argumento minimizador de la suma de los errores cuadrados sobre el espacio de las primeras k autofunciones del operador de covarianza empírico asociado a la muestra $x_1, \ldots, x_n$. La consistencia del mismo la obtienen bajo hipótesis sobre el distanciamiento entre los autovalores asociados al operador de covarianza y muestran tanto en probabilidad como casi seguramente

$$\|\widehat{b}_{pca} - b\|^2 \to 0.$$

Muchos resultados se han obtenido sobre el estimador de componentes principales y variaciones del mismo. Por ejemplo, en Cai y Hall (2006) se obtienen tasas para el error medido por la Γ-Norma, es decir $\|\widehat{b}_{pca} - b\|_\Gamma^2$, y se prueba su optimalidad sobre una amplia clase de funciones. Las hipótesis sobre el distanciamiento de los autovalores se mantienen con variaciones. En general se asume que estos se comportan geométricamente, es decir si λ_n son los autovalores asociados al operador de covarianza, existen $r > 1$ y $C > 0$ tales que para todo n

$$\frac{1}{C} n^{-r} \leq \lambda_n \leq C n^{-r}.$$

También se asumen hipótesis sobre la suavidad de b mediante pertenencia a clases de funciones definidas en la Ecuación 30. Con hipótesis similares, en Hall (2007) obtienen tasas para $\|\widehat{b}_{pca} - b\|^2$. En Cardot (2007) se obtiene un teorema central del límite para el error de predicción entre el estimador y la función de peso proyectada sobre el espacio estimado. Esto permite obtener un test asintótico para contrastar la nulidad de la función de peso. En ese mismo trabajo se argumenta por que no puede existir un resultado tipo teorema central del límite para la norma L_2. En González-Manteiga (2011) proponen un procedimiento bootstrap que permite calcular intervalos de confianza puntuales y para el cual prueban su validez asintótica. Para obtener estos resultados se asumen condiciones adicionales sobre la función de regresión y proceso que genera las covariables.

Otro enfoque para la estimación consiste en minimizar la suma de errores al cuadrado penalizados. Para esto se suele utilizar un espacio de funciones lo suficientemente amplio sobre el cual realizar la minimización controlando la variabilidad de la estimación mediante una penalización. En Cai y Hall (2006) y Hall y Horowitz (2007) se prueban tasas en Γ-Norma y Norma L_2 para un estimador definido a partir de una penalización Ridge sobre la norma de la función estimada. En Yuan y Cai (2010) proponen una metodología por núcleos de reproducción de Hilbert para la que obtienen simultáneamente resultados para la convergencia en Γ-Norma y norma L_2. Así mismo en Cardot et al. (2007) abordan el problema de estimación de la función de peso cuando existen errores de medición. Es decir, cuando se cuenta con observaciones ruidosas de las covariables sobre una grilla . La minimización se efectua sobre un espacio de Sobolev cuya solución se encuentra en un espacio de splines con la mismos quiebres que la grilla de las covariables. Obtienen algunos resultados para muestras finitas asumiendo que existe una constante que acota la norma del proceso con probabilidad uno. En Crambes et al. (2009) se desprenden de esta última hipótesis obteniendo fuertes resultados asintóticos y para muestras finitas respecto a la Γ-Norma.

En los estimadores penalizados las hipótesis sobre distanciamiento de los autovalores ya no suelen ser necesarias para obtener resultados sobre la Γ-Norma. Sin embargo son habituales condiciones

sobre la velocidad de decrecimiento de los autovalores. Muchas veces se pide que su decrecimiento sea geométrico, es decir, que existan $r > 1$ y $C > 0$ tal que

$$\lambda_n \leq Cn^{-r}.$$

También son usuales condiciones sobre la suavidad del proceso o cantidad de derivadas continuas de las covariables y la función de peso. Las hipótesis y las tasas varían si se estudia el error en norma L_2 o en Γ-Norma.

Las propiedades de estimadores penalizados sobre espacios de dimensión finitas fueron estudiadas en Cardot et al. (2003) donde se presenta un estimador por splines penalizado. Acá la dimensión del espacio es un valor prefijado. Para probar su consistencia en norma L_2 se asume que existe una constante que acota la norma del proceso de las covariables con probabilidad uno. También se obtienen tasas para la convergencia que resultan del orden $n^{-\alpha}$ con $\alpha < 0{,}5$. Estas dependen de un parámetro de continuidad de la función de peso. En Apanasovich (2008) estudian el error en Γ-Norma para estimadores basados en un espacio de dimensión finita y una penalización Ridge obteniendo resultados para muestras finitas suponiendo que existe una constante que acota la norma L_2 del proceso con probabilidad 1. Este tipo de estimadores son los que consideramos en este trabajo. De hecho, estudiamos una familia aún más amplia que la estudiada en Apanasovich (2008) ya que trabajaremos con penalizaciones más generales. Utilizamos también hipótesis mucho menos restrictivas.

Es importante mencionar que existen otros estimadores y desarrollos del modelo de regresión lineal funcional. Por ejemplo, Maronna y Yohai (2013) plantea un MM estimador para la función de peso, James (2009) presenta un estimador interpretable de la función de peso introduciendo una penalización Lasso sobre las derivadas. Yao et al. (2005) estudia la regresión lineal con datos longitudinales, James (2002) y Müller y Stadtmüller (2005) estudian el modelo lineal funcional generalizado. Hall y Hooker (2016) estudia el Modelo lineal Truncado. Una revisión más extensa sobre el modelo lineal funcional se puede encontrar en Morris (2015) así como resúmenes sobre el área de datos funcionales y regresión funcional en general sugerimos los trabajos de Horváth y Kokoszka (2012), Cuevas (2014), Goia y Vieu (2016) y Wang et al. (2016).

A pesar de que en estimación no paramétrica la estimación bajo restricciones es un área de fuerte desarrollo no encontramos en la literatura de datos funcionales trabajos sobre estimación del modelo lineal funcional con restricciones de forma.

3.1. Modelos de regresión generales

Consideramos un modelo de regresión general donde observamos replicaciones de pares (x, Y) y modelamos la relación entre x e Y como

$$Y = g(x) + \varepsilon, \tag{31}$$

donde g es una función de las covariables x y toma valores reales. Si bien trabajamos con covariables en un espacio de Hilbert separable, los resultados de esta sección son válidos para cualquier espacio de Banach. Asumiremos la siguiente condición respecto a la relación entre los errores y las covariables

(C1)

$$\mathtt{E}\,(\varepsilon \,|\, x) = 0. \tag{32}$$

Bajo esta condición $g(x) = \mathtt{E}(Y \,|\, x)$ y minimiza $\mathtt{E}((Y - f(x))^2)$ sobre el espacio de todas las funciones con esperanza cuadrada finita.

Sea $\mathcal{H}$ un conjunto de funciones con esperanza cuadrada finita al que g no pertenece necesariamente y $\mathrm{PEN} : \mathcal{H} \to \mathbb{R}_{\geq 0}$ una función de penalización. Si existe, llamemos h_λ a un elemento de $\mathcal{H}$ que minimiza

$$\mathtt{E}(Y - h(x))^2 + \lambda\,\mathrm{PEN}(h). \tag{33}$$

Bajo la condición 32 resulta que $\mathtt{Cov}(\varepsilon, g(x) - h_\lambda(x)) = 0$ y por lo tanto también será un elemento minimizador de

$$\mathtt{E}(g(x) - h(x))^2 + \lambda\,\mathrm{PEN}(h). \tag{34}$$

Si bien el espacio $\mathcal{H}$ puede no ser apropiado para explicar g con suficiente nivel de detalle, mostraremos que si se cumplen ciertas condiciones, tendrá una interesante propiedad respecto al error del modelo explicado por h_λ. Definimos este error como

$$\varepsilon_\lambda := Y - h_\lambda(x) = \varepsilon + g(x) - h_\lambda(x).$$

Le pediremos al espacio $\mathcal{H}$ la siguiente condición:

(C2) Existe un intervalo cerrado I con interior no vacío que contiene al 1 tal que para todo $h \in \mathcal{H}$, $c_1 \in \mathbb{R}$ y $c_2 \in I$

$$c_1 + c_2 h \in \mathcal{H}. \tag{35}$$

La función de penalización la asumiremos no negativa, positiva homogénea de grado $q \in \{0, 1, 2\}$ e invariante por corrimientos. Es decir, cumple con las siguientes propiedades para todo $h \in \mathcal{H}$.

(C3)

$$\mathrm{PEN}(c_2 h + c_1) = c_1^q\,\mathrm{PEN}(h) \quad \forall c_1 \in \mathbb{R} \text{ y } c_2 \in I, \tag{36}$$

$$\mathrm{PEN}(h) \geq 0, \tag{37}$$

$$\mathrm{PEN}(h) = 0 \iff h \equiv c \in \mathbb{R}. \tag{38}$$

Observación 3.1 *Si consideramos el modelo lineal multivariado con dimensión fija con $x \in \mathbb{R}^p$ y $\mathcal{H} = \{h(x) = a_0 + \sum_{j=1}^p a_j x_j \quad a_i \in \mathbb{R}\}$ el espacio satisface (C2). Además si consideramos las penalizaciones Signo, definida en Chebi (2019), Lasso y Ridge estas satisfacen 36 para $q = 0, 1, 2$, respectivamente. La condición 38 se satisface si no se penaliza el intercept.*

El siguiente Teorema es una generalización del resultado obtenido en Hall y Hooker (2016) para el estimador de cuadrados mínimos sin penalizar.

Teorema 3.2 *Bajo el modelo 31 con la condición (C1) sean $\mathcal{H}$ y* PEN *tales que satisfacen (C2) y (C3) resulta*

$$\mathtt{E}\,(h_\lambda(x)) = \mathtt{E}(g(x)), \tag{39}$$

$$\mathtt{E}(h_\lambda^2(x)) = \mathtt{E}(g(x) h_\lambda(x)) - \lambda a_q\,\mathrm{PEN}(h_\lambda), \tag{40}$$

con $a_q = 0, 1/2$ o 1 si $q = 0, 1$ o 2, respectivamente.

Dem: 9.1

El siguiente corolario nos dice que la covarianza entre la parte del modelo que explica h_λ y el error está controlada por la cantidad de penalización.

Corolario 3.3 *Bajo el modelo 31, sean $\mathcal{H}$ y* PEN *tales que satisfacen (C2) y (C3). Si se satisface (C1) y existe h_0 un elemento minimizador de 33 con $\lambda = 0$ entonces*

$$0 \leq \texttt{Cov}(\varepsilon_\lambda, h_\lambda(x)) \leq a_q(h_0)\lambda \quad \forall \lambda \geq 0.$$

Dem: 9.2

3.2. Modelo de Regresión Lineal con restricciones

En esta sección consideramos el Modelo de Regresión Lineal con covariable x en un espacio de Hilbert separable $\mathbb{H}$. La respuesta Y se relaciona con la covariable x mediante la ecuación

$$Y = \alpha + \langle b, x \rangle + \varepsilon, \tag{41}$$

donde $b \in \mathcal{C} \subseteq \mathbb{H}$ un conjunto convexo, $\alpha \in \mathbb{R}$ y ε es una variable aleatoria con esperanza 0, varianza σ^2 e independiente de x.

Si asumimos que la función de regresión b vive en un subconjunto convexo $\mathcal{C} \subseteq H$ decimos que estamos ante un Modelo de Regresión Lineal con Restricciones.

Nuestro interés se centra en la estimación de α y b a partir de una muestra (x_i, Y_i) con $1 \leq i \leq n$ que siguen el modelo con $\texttt{Cov}(\varepsilon_i, \varepsilon_j) = 0$ para $i \neq j$ donde las covariables x_i son estacionarias. Es decir, todas fueron generadas con la misma media m y el mismo operador de covarianza Γ.

A pesar de que la estimación de b en el modelo sin restricciones es adecuada para el modelo con restricciones, estudiamos estimadores que hacen uso de esta información al pertenecer a la clausura de $\mathcal{C}$

Recordamos que un conjunto $\mathcal{C}$ se dice convexo si dados $b_1, b_2 \in \mathcal{C}$ para todo $t \in [0, 1]$

$$tb_1 + (1 - t)b_2 \in \mathcal{C}.$$

3.2.1. Definición de los estimadores

Sean $H \subseteq \mathbb{H}$ un subespacio de dimensión finita $p \in \mathbb{N}$, un operador lineal $D : H \to \mathbb{H}$ que induce la función de penalización $\text{Pen}(u) = \|D(u)\|^2$ y un parámetro de suavizado $\rho \in \mathbb{R}_{\geq 0}$. Si D es inyectivo, esta penalización satisface la condición (C3) con $q = 2$.

Los estimadores de cuadrados mínimos penalizados $(\widehat{\alpha}, \widehat{b})$ pertenecen al conjunto

$$\underset{(\alpha,u)\in\mathbb{R}\times H}{\text{argmin}} \frac{1}{n}\sum_{i=1}^{n}(Y_i - \alpha - \langle x_i\,,\, u\rangle)^2 + \rho\,\text{Pen}(u). \tag{42}$$

Como el conjunto de argumentos minimizadores podría tener más de un elemento debemos elegir uno entre ellos. Para esto, observemos que si (a, u) es un argumento minimizador de 42 necesariamente $a = \bar{Y} - \langle \overline{x}\,,\, u\rangle$. Esto nos basta para definir los estimadores de cuadrados mínimos como $\widehat{\alpha} = \bar{Y} - \left\langle \overline{x}\,,\, \widehat{b}\right\rangle$ y $\widehat{b}$ el elemento de mínima norma en el conjunto de

$$\operatorname*{argmin}_{u\in H} \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y} - \langle x_i - \bar{x}\,,\, u\rangle)^2 + \rho\,\mathrm{Pen}(u). \tag{43}$$

Para el modelo restringido consideramos un conjunto $E \subseteq \overline{\mathcal{C}\cap H}$ convexo y cerrado. Definimos los estimadores de cuadrados mínimos penalizados restringido como $\widetilde{\alpha} = \bar{Y} - \left\langle \overline{x}\,,\, \widetilde{b}\right\rangle$ y $\widetilde{b}$ como el elemento de mínima norma del conjunto

$$\operatorname*{argmin}_{u\in E} \sum_{i=1}^{n}(Y_i - \bar{Y} - \langle u\,,\, x_i - \overline{x}\rangle)^2 + \rho\,\mathrm{Pen}(u). \tag{44}$$

Evidentemente $(\widetilde{\alpha}, \widetilde{b})$ serán argumentos minimizadores de 42 al cambiar la restricción $\mathbb{R}\times H$ por $\mathbb{R}\times E$.

Observación 3.4 *(Buena definición)*

Ambos estimadores están bien definidos ya que la existencia y unicidad de un elemento minimizador de norma mínima está garantizada por la convexidad de la función objetivo y del espacio de argumentos. Esto se debe a que tanto H y E son conjuntos cerrados y convexos y la función objetivo es convexa. Por lo tanto el conjunto de minimizadores es un conjunto convexo y no vacío lo que garantiza la existencia de un único elemento de norma mínima. Para más información sobre análisis convexo se puede consultar Bertsekas (2009).

Lema 3.5 *El estimador restringido $\widetilde{b}$ resulta la proyección de $\widehat{b}$ sobre el espacio E bajo la métrica inducida por la muestra y la penalización. Más precisamente,*

$$\widetilde{b} \in \operatorname*{argmin}_{u\in E} \|\widehat{b} - u\|^2_{\Gamma_n} + \rho\,\mathrm{Pen}(\widehat{b} - u).$$

Dem : 9.3

La siguiente proposición nos dice que la Γ-Norma define el término principal del error cuadrático medio de predicción de los estimadores propuestos.

Proposición 3.6 *Sea $(x_1, Y_1), \ldots, (x_n, Y_n)$ una muestra aleatoria del modelo 41. Si se satisface la condición (A2), los errores no están correlacionados, $b^\star$ es un $\mathbb{H}$ elemento aleatorio y x_{n+1} es un $\mathbb{H}$ elemento aleatorio independiente de $b^\star$ con una ley de distribución con la misma media m y operador de covarianza Γ que la muestra, entonces para $\alpha^\star = \bar{Y} - \langle b^\star\,,\, \overline{x}\rangle$ resulta*

$$\mathtt{E}\left((\alpha + \langle b\,,\, x_{n+1}\rangle - (\alpha^\star + \langle b^\star\,,\, x_{n+1}\rangle))^2 \,\middle|\, b^\star\right) = \mathcal{O}_P\left(\|b - b^\star\|^2_\Gamma + \mathtt{m}_n\|b - b^\star\|^2\right). \tag{45}$$

Dem 9.4

3.2.2. Cálculo de los estimadores

En este apartado derivamos una fórmula explícita para el estimador del modelo sin restricciones. Con este fin llamamos $\tilde{x}_i = n^{-1/2}(x_i - \overline{x})$ e $\tilde{y}_i = n^{-1/2}(Y_i - \bar{Y})$. El vector $\tilde{y} \in \mathbb{R}^n$ satisface que $(\tilde{y})_i = \tilde{y}_i$. Sea $B = \{\delta_1 \dots \delta_p\}$ una base de H, definimos $X \in \mathbb{R}^{n\times p}$ y $W \in \mathbb{R}^{p\times p}$ de manera que

$$(X)_{ij} = \langle \tilde{x}_i \, , \, \delta_j \rangle \qquad (W)_{kj} = \langle D(\delta_k) \, , \, D(\delta_j) \rangle \, . \tag{46}$$

De esta manera tenemos que para $u \in H$ si $|u|$ es la representación de u en la base B, vale que

$$\begin{aligned} \langle u \, , \, x_i - \overline{x} \rangle &= \sqrt{n}(X|u|)_i, \\ \|u\|^2_{\Gamma_n} &= |u|^t X^t X |u|, \\ \mathrm{Pen}(u) &= |u|^t W |u|. \end{aligned} \tag{47}$$

Con esta notación, el problema de minimización 43 equivale a

$$|\widehat{b}| \in \operatorname*{argmin}_{v\in\mathbb{R}^p} \frac{1}{2} v^t (X^t X + \rho W) v - v^t X^t \tilde{y}.$$

Una condición necesaria y suficiente para el argumento minimizador es satisfacer el siguiente sistema de ecuaciones lineales cuyo conjunto de soluciones forma un espacio afín no vacío,

$$(X^t X + \rho W) v = X^t \tilde{y}. \tag{48}$$

Si $X^t X + \rho W$ es inversible, la solución es única y se obtiene como

$$|\widehat{b}| = (X^t X + \rho W)^{-1} X^t \tilde{y}.$$

En el caso de no ser inversible, la solución que buscamos es aquella cuyos coeficientes representen el elemento de H de mínima norma. Estos corresponden a los coeficientes cuya norma es mínima en $\mathbb{R}^p$, cuando B es una base ortonormal y se obtiene utilizando la inversa generalizada de Moore-Penrose que notamos con $\dagger$.

$$|\widehat{b}| = (X^t X + \rho W)^{\dagger} X^t \tilde{y}.$$

Para obtener una expresión del estimador restringido consideramos el isomorfismo $T : \mathbb{R}^p \to H$ tal que $|T(v)|_B = v$. El problema de minimización de la Ecuación 44 equivale al siguiente problema de optimización convexa restringida

$$\begin{aligned} |\widetilde{b}| \in \operatorname*{argmin}_{v\in\mathbb{R}^p} & \frac{1}{2} v^t (X^t X + \rho W) v - v^t X^t \tilde{y} \\ r.a. \quad & T(v) \in E. \end{aligned} \tag{49}$$

3.3. Propiedades para tamaño de muestra fijo

Los siguientes teoremas nos dan cotas para la norma empírica y la norma de los estimadores en muestras finitas. A partir de estas y haciendo uso del Teorema 2.14 podemos obtener tasas en Γ-Norma y en norma cuadrada.

Hacemos uso de las siguientes condiciones para una muestra aleatoria $(x_1, Y_1), \dots, (x_n, Y_n)$ del Modelo Lineal Funcional definido por la Ecuación 41.

Condiciones sobre las covariables y los errores

(M1) Cada x_i con $1 \leq i \leq n$ tiene media m y operador de covarianza Γ.

(M2) Los errores ε_i con $1 \leq i \leq n$ tienen media cero y varianza σ^2 además $\mathtt{Cov}(\varepsilon_i, \varepsilon_j) = 0$ si $i \neq j$.

(M3) Los errores son independientes de las covariables.

En lo que sigue consideramos $D : H \to \mathbb{H}$ una transformación lineal y penalización inducida $\mathrm{Pen}(u) = \|Du\|^2$. Llamamos $0 < \tau_1 \leq \tau_2 \leq \ldots \leq \tau_p$ a los autovalores de D^*D asociados a la matriz W definida en la Ecuación 46. Estos tienen la propiedad enunciada por el Teorema 1.19. Además $\mathtt{s}_k$ queda definido como en el Teorema 2.10.

Sea $G \subseteq \mathbb{H}$ un subespacio cerrado. Si $f \in \mathbb{H}$ llamaos $\Pi_G(f)$ a la proyección ortogonal de f sobre el espacio G. En ocasiones notaremos $f_G = \Pi_G(f)$.

La esperanza condicional a la muestra la notamos por

$$\mathtt{E}_\varepsilon(\cdot) = \mathtt{E}\,(\,\cdot \mid x_1, \ldots, x_n).$$

3.3.1. Estimador de cuadrados mínimos penalizados

Teorema 3.7 *Sea $(x_1, Y_1), \ldots, (x_n, Y_n)$ una muestra aleatoria del Modelo 41 que satisface las condiciones (M1) a (M3), entonces para cualquier $u \in H$*

$$\mathtt{E}\left(\|\mathtt{E}_\varepsilon(\widehat{b}) - b\|^2_{\Gamma_n}\right) \leq 4\mathtt{s}_0\|b - u\|^2 + \rho\,\mathrm{Pen}(u), \tag{50}$$

$$\mathtt{E}\left(\|\widehat{b} - \mathtt{E}_\varepsilon(\widehat{b})\|^2_{\Gamma_n}\right) \leq \sigma^2\frac{p}{n}. \tag{51}$$

Más aún, si $\rho > 0$ para cualquier $1 \leq k, l \leq p$ resulta que

$$\mathtt{E}\left(\|\widehat{b} - \mathtt{E}_\varepsilon(\widehat{b})\|^2_{\Gamma_n}\right) \leq \frac{\sigma^2}{n}\left(k + l + \frac{\mathtt{s}_k}{\rho\tau_l}\right). \tag{52}$$

Dem: 9.8

Corolario 3.8 *Sea $(x_1, Y_1), \ldots, (x_n, Y_n)$ una muestra aleatoria del Modelo 41 que satisface las condiciones (M1) a (M3), entonces para cualquier $u \in H$*

$$\mathtt{E}\left(\|\widehat{b} - b\|^2_{\Gamma_n} + \rho\,\mathrm{Pen}(\widehat{b})\right) \leq 4\mathtt{s}_0\|b - u\|^2 + \rho\,\mathrm{Pen}(u) + \sigma^2\frac{p}{n}. \tag{53}$$

Más aún, si $\rho > 0$ para cualquier $1 \leq k, l \leq p$ resulta que

$$\mathtt{E}\left(\|\widehat{b} - b\|^2_{\Gamma_n} + \rho\,\mathrm{Pen}(\widehat{b})\right) \leq 4\mathtt{s}_0\|b - u\|^2 + \rho\,\mathrm{Pen}(u) + \frac{\sigma^2}{n}\left(k + l + \frac{\mathtt{s}_k}{\rho\tau_l}\right). \tag{54}$$

Dem: 9.9

Consideramos la siguiente condición adicional

(M4) Los errores satisfacen que para cualquier matriz $A \in \mathbb{R}^{n\times n}$ simétrica y definida positiva cumplen que $\mathtt{Var}(\widetilde{\varepsilon}^t A\widetilde{\varepsilon}) \leq \mathrm{Tr}(A\Sigma A\Sigma)$.

Esta hipótesis se cumple bajo normalidad

Proposición 3.9 *Si $\widetilde{\varepsilon} \sim \mathcal{N}(0,\Sigma)$ entonces para cualquier matriz $A \in \mathbb{R}^{n\times n}$ vale que*

$$\mathtt{Var}(\widetilde{\varepsilon}^t A\widetilde{\varepsilon}) = Tr(A\Sigma A\Sigma).$$

Dem: Ver Teorema 5.2c en Rencher (2008).

El siguiente Teorema de crucial importancia establece las condiciones que se deben cumplir para controlar la esperanza de la norma cuadrada del estimador.

Teorema 3.10 *Sea $(x_1,Y_1),\ldots,(x_n,Y_n)$ una muestra aleatoria del Modelo 41 que satisface las condiciones (M1) a (M3), si $b \in Nu(\Gamma)^{\perp}$ y $\rho > 0$ resulta que*

$$\mathtt{E}\left(\|\widehat{b}\|^2\right) \leq \left(\|b\| + \sqrt{\frac{4\mathtt{s}_0}{\rho\tau_1}}\|b - b_H\|\right)^2 + \frac{\sigma^2}{n\rho}\sum_{j=1}^{p}\tau_j^{-1}. \tag{55}$$

Más aún, si los errores también satisfacen la hipótesis (M4) y x un elemento aleatorio con la misma distribución de los x_i entonces

$$\mathtt{E}\left(\|\widehat{b}\|^4\right) \leq 64\left(\|b\|^4 + \left(\frac{\sigma^2}{n\rho}\sum_{j=1}^{p}\tau_j^{-1}\right)^2 + 16\frac{\|b-b_H\|^4}{\rho^2\tau_1^2}\mathtt{E}\left(\|x-m\|^4\right)\right). \tag{56}$$

De manera adicional, si $b \notin Nu(\Gamma)^{\perp}$ las desigualdades anteriores valen reemplazando b por $\Pi_{Nu(\Gamma)^{\perp}}(b)$.

Dem: 9.10

Observación 3.11 *En la demostración de la Desigualdad 55 usamos fuertemente la hipótesis de que ρW es una matriz inversible. Si la hipótesis no se cumple el resultado valdría igual pero reemplazando $(\rho\tau_1)^{-1}$ por $\tau_{min}(X^tX)^{-1}$. Como los autovalores de X^tX tienden a cero cuando $p \to +\infty$ no será posible controlar el término al hacer un análisis asintótico.*

Observación 3.12 *Un resultado similar al de la Ecuación 56 puede obtenerse relajando la hipótesis (M4) a la existencia de una constante $C > 0$ tal que*

$$\mathtt{Var}(\widetilde{\varepsilon}^t A\widetilde{\varepsilon}) = C\,Tr(A\Sigma A\Sigma).$$

El valor 64 debe ser reemplazado por una constante que depende de C.

3.3.2. Estimador restringido de cuadrados mínimos penalizados

El siguiente teorema nos da desigualdades no aleatorias que nos servirán para probar tasas en Γ-norma para el estimador restringido y valen para una muestra $(x_1, Y_1), \dots (x_n, Y_n)$ del modelo sin asumir ninguna condición sobre esta. Estas desigualdades son determinísticas y nos permiten comparar las tasas en norma empírica de la estimación sin restricciones con la norma empírica de la restringida.

Teorema 3.13 *Si $E \subseteq H$ es un conjunto cerrado y convexo,* $\mathrm{Pen}(u) = 0$ *si y solo si* $u = 0$*, entonces para todo $u \in E$ vale que*

$$\|\widetilde{b} - b\|_{\Gamma_n}^2 \leq 4\|\widehat{b} - b\|_{\Gamma_n}^2 + 6\|b - u\|_{\Gamma_n}^2 + 4\rho\,\mathrm{Pen}(\widehat{b} - u), \tag{57}$$

$$\|\widetilde{b}\|^2 \leq 2\|\widehat{b}\|^2 + 2(\tau_1\rho)^{-1}\left(\|\widehat{b} - u\|_{\Gamma_n}^2 + \rho\,\mathrm{Pen}(\widehat{b} - u)\right). \tag{58}$$

Dem: 9.11

La siguiente observación es consecuencia inmediata del Teorema 3.13 y el Corolario 3.8.

Observación 3.14 *Bajo las condiciones del Teorema 3.13 y del Corolario 3.8, dados $l, k \leq p$ llamamos $c_{n\,k\,l} = \sigma^2 \min\left\{\frac{p}{n}, \frac{1}{n}(k + l + \frac{\mathtt{s}_k}{\rho\tau_l})\right\}$, resulta que para todo $u \in E$ valen las siguientes desigualdades*

$$\mathtt{E}\left(\|\widetilde{b} - b\|_{\Gamma_n}^2\right) \leq 56\mathtt{s}_0\|b - u\|^2 + 16\rho\,\mathrm{Pen}(u) + 8c_{n\,k\,l}, \tag{59}$$

$$\sup_{n\in\mathbb{N}} \mathtt{E}\left(\|\widetilde{b}\|^2\right) < +\infty. \tag{60}$$

3.4. Comportamiento asintótico de los estimadores

En esta sección estudiamos el comportamiento asintótico de los estimadores. Para esto consideramos H_n una sucesión de espacios de dimensión finita p_n y $D_n : H_n \to \mathbb{H}$ una sucesión de operadores lineales que inducen una sucesión de penalizaciones $\mathrm{Pen}_n(u) = \|D_n(u)\|^2$.

Decimos que la sucesión de penalizaciones inducidas es acotada si existe una constante $C_1 > 0$ tal que para todo $n \in \mathbb{N}$

$$\|D_n u\|^2 \leq C_1\|u\|^2 \quad \forall u \in H_n.$$

Decimos que la sucesión de penalizaciones es coerciva si existe una constante $C_2 > 0$ tal que para todo $n \in \mathbb{N}$

$$\|D_n u\|^2 \geq C_2\|u\|^2 \quad \forall u \in H_n.$$

Observación 3.15 *Tomando D_n como la identidad se induce a una sucesión de penalizaciones acotada y coerciva.*

La siguiente proposición nos permite obtener las tasas y consistencias bajo las distintas normas. Las Figuras 4 y 5 ayudan a visualizar las distintas posibilidades.

Proposición 3.16 *Sea* $(x_1, Y_1), \ldots, (x_n, Y_n)$ *una muestra aleatoria del Modelo 41 que satisface las condiciones (M1) a (M3). Sea* H_n *una sucesión de subespacios de dimensión* p_n *tales que* $\|b - b_{H_n}\|^2 \to 0$ *y sea* Pen_n *una sucesión de penalizaciones acotada y coerciva. Si se toma* ρ_n *de manera que* $\max\{\|b - b_{H_n}\|^2, \frac{p_n}{n}\} = \mathcal{O}(\rho_n)$ *resulta*

$$\mathtt{E}\left(\|\widehat{b}_n - b\|^2_{\Gamma_n}\right) = \mathcal{O}(\rho_n),$$
$$\sup_{n\in\mathbb{N}} \mathtt{E}\left(\|\widehat{b}_n\|^2\right) < +\infty.$$

Si en cambio ρ_n *es tal que* $\max\{\|b - b_{H_n}\|^2, \frac{p_n}{n}\} = o(\rho_n)$ *entonces*

$$\mathtt{E}\left(\|\widehat{b}_n - b\|^2_{\Gamma_n}\right) = \mathcal{O}(\rho_n),$$
$$\limsup_{n\in\mathbb{N}} \mathtt{E}\left(\|\widehat{b}_n\|^2\right) \leq \|\Pi_{Nu(\Gamma)^\perp}(b)\|^2.$$

En ambas situaciones, si los errores satisfacen (M4) entonces

$$\sup_{n\in\mathbb{N}} \mathtt{E}\left(\|\widehat{b}_n\|^4\right) < +\infty.$$

Dem: 9.12

Observación 3.17 *Los casos en que la sucesión de penalizaciones no es acotada o coerciva deben ser estudiados de manera particular. Que esté acotada es importante para la Ecuación 50 ya que* $\mathrm{Pen}(b_{H_n})$ *no está necesariamente acotado y* $\mathtt{E}(\|\widehat{b}_n - b\|^2_{\Gamma_n})$ *podría no tender a cero. La coercividad la necesitamos para poder usar la Desigualdad 55 y controlar el término* $\mathtt{E}(\|\widehat{b}\|^2)$.

Gracias a la proposición anterior podemos construir un estimador consistente en un conjunto muy amplio de funciones sin asumir hipótesis sobre el operador de covarianza.

Observación 3.18 *Si* $\{f_j \quad j \in \mathbb{N}\}$ *es una base ortonormal de* $\mathbb{H}$*, el espacio* $H_p = \langle f_1, \ldots, f_p\rangle$ *satisface*

$$\|b - b_{H_p}\|^2 = \sum_{j=p+1}^{+\infty} \langle b\,,\, f_j\rangle^2\,.$$

Utilizando la penalización $\mathrm{Pen}(u) = \|u\|^2$*, una sucesión de espacios* H_{p_n} *de manera tal que* $p_n = o(n)$ *y una sucesión de parámetros de suavizado* ρ_n *tal que*

$$\max\left\{\sum_{j=p_n+1}^{+\infty} \langle b\,,\, f_j\rangle^2\,, \frac{p_n}{n}\right\} = o(\rho_n),$$

entonces si se satisfacen (A1), (A2), (M1) - (M3), se obtiene

$$\lim_{n\to+\infty} \mathtt{E}\left(\|\widehat{b}_n - \Pi_{Nu(\Gamma)^\perp}(b)\|^2\right) = 0.$$

La siguiente proposición deja explícitas las condiciones bajo las cuales las tasas en Γ-Norma del estimador restringido son menores o iguales a las de la estimación sin restringir.

Proposición 3.19 *Sea H_n una sucesión de subespacios de dimensión p_n y $E_n \subseteq \mathcal{C} \cap H_n$ convexo y cerrado para el cual existe $u_n \in E_n$ tal que $\|b-u_n\|^2 \to 0$. Si $\mathcal{C}$ es un conjunto convexo y la sucesión* Pen_n *es coerciva y acotada entonces con ρ_n tal que* $\text{máx}\left\{\|b-u_n\|^2, \frac{p_n}{n}\right\} = \mathcal{O}(\rho_n)$ *tenemos*

$$\mathtt{E}\left(\|\widetilde{b}-b\|^2_{\Gamma_n}\right) = \mathcal{O}\left(\rho_n\right),$$
$$\mathtt{E}\left(\|\widetilde{b}\|^2\right) < +\infty.$$

Si se satisface que $\|b-u_n\|^2 = \mathcal{O}\left(\|b-b_{H_n}\|^2\right)$ obtenemos

$$\mathtt{E}\left(\|\widetilde{b}-b\|^2_{\Gamma_n}\right) = \mathcal{O}\left(\mathtt{E}\left(\|\widehat{b}-b\|^2_{\Gamma_n}\right)\right).$$

Dem: Consecuencia directa de la Proposición 3.16 y el Teorema 3.13.

Cuando $\mathcal{C}$ es un conjunto acotado podemos quitar la hipótesis de coercividad la cual es necesaria para el estimador sin restricciones ya que el término $\frac{1}{n\rho}\sum_{j=1}^{p}\tau_j^{-1}$ de la Desigualdad 55 no lo podemos controlar.

Proposición 3.20 *Sea H_n una sucesión de subespacios de dimensión p_n y $E_n \subseteq \mathcal{C} \cap H_n$ convexo y cerrado para el cual existe $u_n \in E_n$ tal que $\|b-u_n\|^2 \to 0$. Si $\mathcal{C}$ es un conjunto acotado y la sucesión* Pen_n *es acotada e inducida por una sucesión de operadores D_n inyectivos y acotados, entonces con ρ_n tal que* $\text{máx}\left\{\|b-u_n\|^2, \frac{p_n}{n}\right\} = \mathcal{O}(\rho_n)$

$$\mathtt{E}\left(\|\widetilde{b}_n-b\|^2_{\Gamma_n}\right) = \mathcal{O}(\rho_n),$$
$$\sup_{n\in\mathbb{N}}\mathtt{E}\left(\|\widetilde{b}_n\|^4\right) < +\infty.$$

Dem: 9.13

Observación 3.21 *Si $\mathcal{C}$ es un conjunto convexo y acotado resulta que para todo $r \in \mathbb{N}$ tenemos* $\sup_{n\in\mathbb{N}}\mathtt{E}(\|\widetilde{b}\|^{2r}) < \infty$ *luego si $\rho_n = o(1)$ por las Proposición 2.18 resulta*

$$\lim_{n\to+\infty}\mathtt{E}\left(\|\widetilde{b}_n-b\|^r_{\Gamma}\right) = 0.$$

3.5. Estimación mediante espacios de splines

En esta sección trabajamos con $\mathbb{H} = L_2[a,d]$ donde $\langle f\,,\,g\rangle = \int_a^d f(t)g(t)\,dt$. Llamamos $C^k[a,d]$ al espacio de funciones con k derivadas continuas y acotadas en el intervalo $[a,d]$. Las derivadas en los bordes se definen mediante límites laterales.

Sea $\Pi : a = t_0 < t_1 < \ldots < t_p = d$ una grilla o partición del intervalo $[a,d]$. Llamamos $P^k([a,d],\Pi)$ al conjunto de funciones que restringidas a cada subintervalo (t_j, t_{j+1}) son un polinomio de grado a

lo sumo k. El espacio de splines de orden k queda definido como $S^k(\Pi) = P^{k-1}(\Pi) \cap C^{k-2}[a,d]$. Este espacio tiene dimensión $p+k-1$. Diremos que la grilla es uniforme si $h = t_{i+1} - t_i$ es independiente de i. Nos referiremos como splines cúbicos al espacio con $k = 4$. Notar que las funciones en S^k son $k-1$ veces derivables a trozos.

El siguiente Teorema establece las excelentes propiedades de aproximación de estos espacios.

Teorema 3.22 *Sea Π una partición uniforme de $[a,d]$. Si $k = 3$ o es par, existe un operador de interpolación $\pi : C^k[a,d] \to S^k(\Pi)$ tal que $\pi g = g$ si $g \in S^k(\Pi)$ y*

$$\|(f - \pi f)^{(r)}\|_\infty \leq D_k \|f^{(k)}\|_\infty h^{k-r} \quad 0 \leq r \leq k-1,$$

donde las constantes D_k no dependen del tamaño de la grilla.

Dem: Ver Hall y Meyer (1976) para el caso con k par. Dubeau y Savoie (1996) para el caso con $k = 3$.

3.5.1. Computo del estimador restringido

En esta sección mostramos como calcular el estimador restringido para los casos donde la restricción se define por la positividad o negatividad de una derivada. Ejemplos de estos casos son la monotonía y la convexidad. Sean $\delta_1, \ldots, \delta_p$ una base del espacio de splines $S^k[a,d]$ de orden k con partición uniforme del intervalo $[a,d]$. Sean $t_0 = a < t_1 < \ldots, < t_{p-k} < t_{p-k+1} = b$ son los puntos de la partición. Como $f \in S^k_\pi[a,d]$, por definición $f|_{[t_j, t_{j+1}]}$ es un polinomio de grado $k-1$. Por lo tanto $f^{(k-2)}\big|_{[t_j,t_{j+1}]}$ es una función lineal. Como los extremos de una lineal sobre un intervalo cerrado se realizan en los bordes, es suficiente evaluar en estos puntos para determinar si es positiva en un intervalo. Por lo tanto la positividad de f^{k-2} se determina mediante un sistema de $t-k+2$ inecuaciones lineales. Es decir, si $f = \sum_{j=1}^p b_j \delta_j$ y consideramos la matriz $C \in \mathbb{R}^{(p-k+2)} \times p$ tal que $C_{kj} = \delta_j(t_k)$ tendremos que

$$f^{(k-2)}(t) \geq 0 \, \forall t \in [a,d] \iff C\vec{b} \geq 0.$$

Con esta observación, el problema de minimización 49 resulta un problema de minimización cuadrática con restricciones lineales. Existen numerosas librerías para la resolución de estos problemas. Para la implementación del algoritmo utilizamos el paquete *Quadprog* de R.

3.5.2. Resultados asintóticos para la estimación sin restricciones

En esta sección analizamos las tasas para la estimación con espacios de splines y penalizaciones coercivas pero no necesariamente acotadas. Además usamos las propiedades de aproximación de los splines para obtener tasas explícitas.

Para $k \in \mathbb{N}$ y $\vec{\alpha} \in \mathbb{R}^k_{\geq 0}$ definimos la penalización

$$\text{Pen}_{\vec{\alpha}}(f) = \sum_{r=1}^k \alpha_r \|f^{(r-1)}\|^2. \tag{61}$$

Notemos que esta penalización está definida en $S^k(\Pi)$ cualquiera sea la partición Π. Más aún, si $\alpha_1 > 0$ es decir aparece la norma cuadrada de la función sin derivar, entonces $\mathrm{Pen}(f) \geq \alpha_1 \|f\|^2$.

Observación 3.23 *Sea $(x_1, Y_1), \ldots (x_n, Y_n)$ una muestra aleatoria que satisface las condiciones (M1) a (M3). Supongamos que se satisface que $b \in C^k[a,d]$ con $k = 3$ o k par. Sea S_p^k el espacio de splines con partición uniforme de orden k y dimensión p. El calibre de la grilla se define $h := t_{j+1} - t_j = \frac{1}{p-k}$. Sea π el operador de interpolación definido por el Teorema 3.22 y sea $\vec{\alpha} \in \mathbb{R}^k$ entonces*

$$\|b - \pi b\|^2 + \rho \,\mathrm{Pen}_{\vec{\alpha}}(\pi b) = \mathcal{O}(h^{2k} + \rho).$$

Tomando $p_n = \lfloor n^{1/(2k+1)} \rfloor$ resulta

$$h^{2k} = \mathcal{O}(n^{-2k/(2k+1)}) \quad y \quad \frac{p_n}{n} = \mathcal{O}(n^{-2k/(2k+1)}).$$

Por el Corolario 3.8

$$\mathtt{E}(\|\widehat{b}_n - b\|_{\Gamma_n}^2) = \mathcal{O}\left(n^{-2k/(2k+1)} + \rho_n\right).$$

Dependiendo de ρ_n tendremos distintos resultados. Si $\rho_n \propto n^{-2k/(2k+1)}$ obtenemos

$$\mathtt{E}(\|\widehat{b}_n - b\|_{\Gamma_n}^2) = \mathcal{O}\left(n^{-2k/(2k+1)}\right), \tag{62}$$

$$\limsup_{n \in \mathbb{N}} \mathtt{E}(\|\widehat{b}\|^2) < +\infty. \tag{63}$$

Si ρ_n es tal que $n^{-2k/(2k+1)} = o(\rho_n)$ perderemos un poco de tasas para la convergencia en norma empírica pero a cambio obtenemos que

$$\limsup_{n} \mathtt{E}(\|\widehat{b}\|^2) \leq \|\Pi_{Nu(\Gamma)^\perp}(b)\|^2.$$

Si $\rho_n \propto n^{-k/(2k+1)}$, la sucesión de penalizaciones es coerciva

$$\mathtt{E}(\|\widehat{b}_n - b\|_{\Gamma_n}^2) = \mathcal{O}\left(n^{-k/(2k+1)}\right),$$
$$\mathtt{E}(\|\widehat{b}_n\|^2) \leq \|\Pi_{Nu(\Gamma)^\perp}(b)\|^2 + \mathcal{O}(n^{-k/(2k+1)}).$$

Si además se satisface la condición (M4)

$$\sup_{n \in \mathbb{N}} \mathtt{E}\left(\|\widehat{b}\|^4\right) < +\infty.$$

La siguiente proposición ilustra las tasas que obtenemos en Γ-Norma

Proposición 3.24 *Si se satisfacen las condiciones (A1) y (A2) para cualquiera de las elecciones de ρ_n de la observación anterior tenemos*

$$\|\widehat{b}_n - b\|_{\Gamma}^2 = \mathcal{O}_P\left(\max\left\{\rho_n, \frac{\mathtt{t}_{k_n}}{n}, \frac{\mathtt{s}_{k_n}}{\sqrt{n}}, \mathtt{m}_n\right\}\right).$$

Si vale la condición (A3) con $\mathtt{l}_n = \mathcal{O}(n^{-1/2})$ *resulta*

$$\mathtt{E}\left(\|\widehat{b}_n - b\|_\Gamma^2\right) = \mathcal{O}\left(\max\left\{\rho_n, n^{-1/2}\right\}\right).$$

Si $\rho_n = o(n^{-2k/(2k+1)})$ *y la sucesión de penalizaciones es coerciva obtenemos por la Proposición 2.19*

$$\lim_{n\to+\infty} \mathtt{E}\left(\|\widehat{b}_n - b\|^2\right) = 0.$$

Si $\rho_n \propto n^{-1/2}$ *tenemos que el término* $\mathtt{d}_n$ *del Corolario 2.21 resulta* $o(n^{-1/4})$*. Luego*

$$\mathtt{E}(\|\widehat{b}_n - b\|^2) \leq n^{-1/4}\lambda_{k_n}^{-1/2} + \sum_{r=k_n+1}^{\infty} \langle b\,,\, e_r\rangle^2\,.$$

Por lo que imponiendo condiciones del tipo $\lambda_k \geq Cn^{-1-a}$ *y sobre la velocidad de convergencia a cero de la sucesión* $b_j = \langle b\,,\, e_j\rangle^2$ *podemos obtener tasas para la convergencia* $\|\widehat{b} - b\|^2$*.*

3.5.3. Resultados asintóticos para la estimación restringida

En este apartado llegamos al problema que motiva la primera parte de esta tesis: Estimación restringida a conjuntos convexos de funciones. Los ejemplos que tenemos en mente son de funciones no negativas, funciones crecientes o funciones convexas pero lo expresamos de una forma más general mediante la siguiente definición que da lugar a la *forma* de una función.

Definición 3.25 *Sea* $k \in \mathbb{N}$ *y el operador diferencial* $L_{\vec{a}} : C^k[a,d] \to C[a,d]$ *definido por* $L_{\vec{a}}(f) = \sum_{r=0}^{k-1} a_k f^{(r)}$*. Definimos los conjuntos*

$$\mathcal{L}_{\vec{a}} = \{f \in C^k[a,d] \,:\, L_{\vec{a}}(f)(t) \geq 0\, \forall t \in [a,d]\},$$
$$\mathcal{L}_{\vec{a},0} = \{f \in C^k[a,d] \,:\, L_{\vec{a}}(f)(t) > 0\, \forall t \in [a,d]\}.$$

Es inmediato ver que estos conjuntos son convexos. Además son una generalización de los conjuntos mencionados ya que $\mathcal{L}_{(1,0,0)} \subseteq \mathcal{L}_{(1,0)} \subseteq \mathcal{L}_{(1)}$, son conjuntos de funciones positivas con distintos niveles de suavidad. De la misma manera $\mathcal{L}_{(0,1)} \subseteq \mathcal{L}_{(0,1,0)}$ son conjuntos de funciones estrictamente crecientes y el $\mathcal{L}_{(0,0,1)}$ de funciones convexas.

La siguiente proposición nos dice que si una función está en el interior del conjunto de las restricciones su proyección sobre el espacio de splines también lo está.

Proposición 3.26 *Sea* $k = 3$*, o par y* $\vec{a} \in \mathbb{R}^k$*. Si* $f \in \mathcal{L}_{\vec{a},0}$ *entonces existe* $h_0 > 0$ *tal que si* Π *es una partición uniforme del intervalo* $[a,d]$ *con calibre* $h \leq h_0$ *entonces* $\pi f \in \mathcal{L}_{\vec{a}}$*.*

Dem: 9.14

En caso de que la función esté en el borde del conjunto de restricciones también podemos afirmar que hay un elemento en el espacio de splines que se encuentra cerca y que satisface las restricciones.

Proposición 3.27 *Sea $k = 3$ o par y $\vec{a} \in \mathbb{R}^k$. Sea $f \in \mathcal{L}_{\vec{a}}$ y sea Π una partición uniforme del intervalo $[a, d]$ con calibre h. Existe $g \in E = \mathcal{L}_{\vec{a}} \cap S^k(\Pi)$ tal que para $r_1 = \max_r\{a_r \neq 0\}$ existe una constante $D_{\vec{a}}$ que solo depende de $\vec{a}$ tal que*

$$\|f - g\|^2 \leq D_{\vec{a}}(b - a)^{r_0} h^{2(k-r_1)}.$$

Dem: 9.15

Finalmente obtenemos las tasas para la norma empírica.

Proposición 3.28 *Sea $\vec{a} \in \mathbb{R}^k$ con $k = 3$ o par y S_p^k el espacio de splines de orden k con partición uniforme y dimensión p. Si $b \in \mathcal{L}_{\vec{a},0}$ y $\widetilde{b}_n$ la sucesión de estimadores restringidos definidos en 44 con $H_n = S_{p_n}^k$ y una sucesión de penalizaciones Pen_n acotada y coerciva. Si $p_n = \lfloor n^{1/(2k+1)} \rfloor$ y $\rho_n \propto \mathcal{O}(n^{-2k/(2k+1)})$ obtenemos*

$$\begin{aligned} \mathtt{E}(\|\widetilde{b}_n - b\|_{\Gamma_n}^2) &= \mathcal{O}(n^{-2k/(2k+1)}), \\ \mathtt{E}(\|\widetilde{b}_n - b\|^2) &= \mathcal{O}(1). \end{aligned}$$

Si $b \in \mathcal{L}_{\vec{a}}$ y $r_1 = \max_r\{a_r \neq 0\}$, tomando $p_n = \lfloor n^{1/(2k-2r_1+1)} \rfloor$ y $\rho_n = n^{-2(k-r_1)/(2k-2r_1+1)}$ obtenemos

$$\begin{aligned} \mathtt{E}(\|b - \widetilde{b}\|_{\Gamma_n}^2) &= \mathcal{O}(n^{-2(k-r_1)/(2(k-r_1)+1)}), \\ \mathtt{E}(\|b - \widetilde{b}\|^2) &= \mathcal{O}(1). \end{aligned}$$

Dem: 9.16

4. Modelo Lineal Histórico

El modelo funcional histórico o modelo funcional truncado es un caso particular del modelo de regresión lineal definido en la Ecuación 1. En este asumimos que el dominio de las covariables es el intervalo $I = [0, T]$, que la función de regresión b es continua y que existe un punto del dominio a partir del cual la función es constantemente nula. El mínimo de los valores para los que esto ocurre lo llamamos punto de cambio o punto de truncado y lo notamos θ_0. Es decir,

$$\theta_0 = \min_{\theta \in [0,T]} \{\theta : b(t) = 0 \ \forall t \geq \theta\}.$$

El modelo resulta equivalente a

$$Y = \alpha + \int_0^{\theta_0} b(t)x(t)\, dt + \varepsilon. \tag{64}$$

Nos enfocamos en estimar θ_0 y en hacer uso de esta restricción del modelo para estimar la función de peso.

En Hall y Hooker (2016) se discute la identificabilidad del modelo y se proponen metodologías de estimación de α b y θ_0 de manera conjunta o en dos pasos obteniendo la consistencia para la estimación de θ_0. En Guan et al. (2020) se propone una metodología Bridge para grupos anidados donde se minimiza la suma de errores cuadrados penalizados sobre un espacio de splines. La penalización utilizada induce a obtener coeficientes nulos en los últimos elementos del desarrollo en una base de B-splines lo que fuerza a la función estimada a ser nula a partir de un punto. También prueban resultados de consistencia para la estimación del punto de cambio.

Con una demostración diferente recuperamos los resultados de Hall y Hooker (2016) y obtenemos tasas de convergencia para la estimación de θ_0. También obtenemos las tasas en Γ-Norma del error de estimación de b para el estimador de tres pasos que es constantemente nulo a partir del punto de cambio estimado.

4.1. Estimación en dos pasos del punto de cambio

Para definir el estimador utilizamos la siguiente notación que resulta conveniente para las demostraciones de consistencia y tasas del punto de truncado. Llamamos $Qh(x, Y)$ a la esperanza de $h(x, Y)$ cuando Q es la distribución de (x, Y). En particular si $(x_1, Y_1), \ldots (x_n, Y_n)$ es una muestra aleatoria independiente con distribución P adoptamos P_n para definir la medida empírica. Por lo tanto $P_n(h(x, Y)) = \frac{1}{n}\sum_{i=1}^n h(x_i, Y_i)$. Más aún llamamos $m(t) = Px(t)$ y $\mu = PY = \alpha + \langle b\,,\, m\rangle$ a los valores esperados de Y y x respectivamente. Asumimos además que x_i son independientes de los errores ε_i y que ambos tienen segundo momento finito, esto es $P(\|x\|^2) < +\infty$ y $P(\varepsilon^2) = \sigma^2$.

Si para cada función $f \in L_2[0, T]$ definimos $f_\theta = \mathbb{I}_{[0,\theta]} f$ el Modelo Lineal Histórico puede ser formulado de la siguiente forma

$$Y = PY + \int_0^{\theta_0} b(t)(x(t) - Px(t))\, dt + \varepsilon = PY + \langle b_{\theta_0}\,,\, x - Px\rangle + \varepsilon. \tag{65}$$

Para $\theta \in [0,T]$, $f \in L_2[0,T]$ y Q una medida en $L_2[0,T] \times \mathbb{R}$ definimos

$$\ell(\theta, f, Q) = Q\left(Y - QY - \langle f_\theta\,,\, x - Qx\rangle\right)^2. \tag{66}$$

En lo que sigue estudiamos propiedades de $\ell(\theta, f, P)$ y $\ell(\theta, f, P_n)$.

Observación 4.1 *Usando la Ecuación 65, que ε es independiente de x y que $b_\theta = b$ si $\theta \geq \theta_0$ resulta*

$$\ell(\theta, b, P) = \mathbb{I}_{\theta<\theta_0} P\,\langle b - b_\theta\,,\, x - Px\rangle^2 + P\varepsilon^2. \tag{67}$$

De la observación se desprende que el punto de cambio θ_0 queda caracterizado como el valor mínimo de los argumentos minimizadores de $\ell(\,\cdot\,, b, P)$ siempre que se cumpla la siguiente condición:

Condición (B1) Para todo $\theta < \theta_0$ vale

$$P\,\langle b - b_\theta\,,\, x - Px\rangle^2 > 0.$$

Dicho de otra manera, si vale la condición $(B1)$ tenemos que

$$\ell(\theta, b, P) > \ell(\theta_0, b, P) \quad \text{ si } \theta < \theta_0 \quad \text{ mientras que } \quad \ell(\theta, b, P) = \ell(\theta_0, b, P) \text{ si } \theta \geq \theta_0. \tag{68}$$

Esta caracterización de θ_0 sugiere una estimación por método *plug-in* de dos pasos. Primero obtener $b^\star$ una estimación de b bajo el modelo 1 luego encontrar $\hat{\theta}$ como el valor más chico que minimiza $\ell(\theta, b^\star, P_n)$. Sin embargo esta definición se encuentra con dificultades ya que la función a minimizar suele alcanzar su mínimo en $\theta = T$. Este caso es habitual cuando $b^\star$ se obtiene minimizando alguna variación de $\ell(T, \,\cdot\,, P_n)$ como es el caso de toda la familia de estimadores definidos en la sección 3.2.1. Para evitar este problema le sumamos un término de penalización,

$$PL(\theta, b^\star, P_n) = \ell(\theta, b^\star, P_n) + \omega_n \text{pen}(\theta),$$

donde $\omega_n \geq 0$ y $\text{pen} : [0,T] \to \mathbb{R}_{\geq 0}$ es una función estrictamente creciente. Ejemplos de posibles funciones son $\text{pen}(\theta) = \theta$ o $\text{pen}(\theta) = \theta^2$. Definimos entonces los estimadores

$$\hat{\theta} = \underset{\theta \in [0,T]}{\text{argmin}}\, PL(\theta, b^\star, P_n), \tag{69}$$

esto quiere decir que $\hat{\theta}$ satisface

$$PL(\hat{\theta}, b^\star, P_n) \leq PL(\theta, b^\star, P_n), \quad \forall \theta \in [0,T].$$

Como estimador inicial es posible usar cualquiera de los que estudiamos para el modelo lineal funcional.

4.2. Estimación del modelo con un punto de truncado

Una vez que obtenida una estimación de $\hat{\theta}_0$ nos interesa estimar la función de peso de manera que sea nula a partir del punto de truncado estimado. Uno quisiera truncar las covariables y la función

de peso y plantear el modelo lineal sobre el intervalo $[0, \hat{\theta}]$. Esto resulta equivalente a estimar la función de peso sobre el espacio de funciones que es constantemente cero a partir de $\hat{\theta}_0$.

Con esta idea en mente, sea $H \subseteq L_2[0,1]$ un subespacio de dimensión p, para cada $\theta \in [0,T]$ consideramos el espacio $H_\theta \subseteq L_2[0,T]$ donde $f \in H_\theta$ si existe $g \in H$ tal que

$$f(t) = \begin{cases} \sqrt{\theta} g(\frac{t}{\theta}) & \text{si } 0 \leq t \leq \theta, \\ 0 & \text{si } \theta < t \leq T. \end{cases} \tag{70}$$

Por comodidad en la notación también definimos los espacios $\tilde{H}_\theta \subseteq L_2[0,\theta]$ donde $\tilde{f} \in \tilde{H}_\theta$ si existe $g \in H$ tal que

$$\tilde{f}(t) = \sqrt{\theta} g(\frac{t}{\theta}).$$

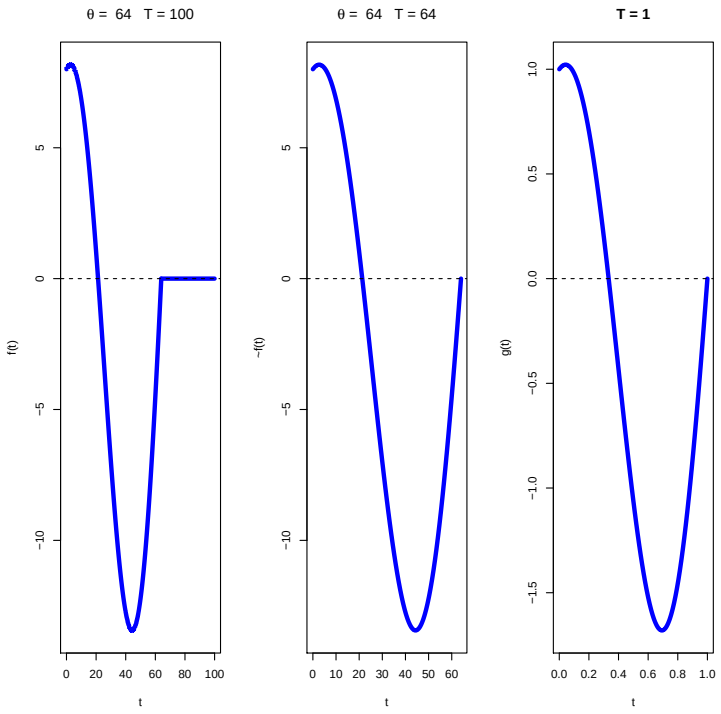

Figura 6: Ejemplo de funciones f, $\widetilde{f}$ y g en sus respectivos espacios con $\theta = 64$ y $T = 100$.

Claramente los espacios H_θ y $\tilde{H}_\theta$ son subespacios de dimensión p. Más aún, si $\{g_1, \ldots, g_p\}$ son una base ortonormal de H también lo serán $\{f_1, \ldots, f_p\}$ y $\{\tilde{f}_1, \ldots \tilde{f}_p\}$ para sus respectivos espacios.

Para cada $\theta \in [0,T]$ definimos

$$\widehat{b}[\theta] = \arg \min_{u \in H_\theta} \sum_{i=1}^{n} (Y_i - \bar{Y} - \langle u\,,\, x_i - \overline{x} \rangle)^2 + \rho\, \mathrm{Pen}(u). \tag{71}$$

En caso de que asumamos que $b \in \mathcal{C}$ un conjunto convexo de $L_2[0,T]$, si $E_\theta \subseteq \overline{\mathcal{C} \cap H_\theta}$ es un conjunto convexo y cerrado definimos

$$\widetilde{b}[\theta] = \arg\min_{u \in E_\theta} \sum_{i=1}^{n} (Y_i - \bar{Y} - \langle u\,,\, x_i - \overline{x} \rangle)^2 + \rho\, \mathrm{Pen}(u). \tag{72}$$

Ambos estimadores están bien definidos por ser casos particulares de las Ecuaciones 42 y 43 respectivamente. Luego si $\hat{\theta}$ es una estimación de θ_0 tenemos los estimadores de tres pasos definidos como $\widehat{b}[\hat{\theta}]$ o $\widetilde{b}[\hat{\theta}]$ que serán continuos siempre que H sea un espacio de funciones continuas que se anulan en uno.

4.3. Resultados Asintóticos

En esta sección mostramos propiedades asintóticas de los estimadores propuestos y mostramos resultados que hasta nuestro conocimiento son nuevos en la literatura.

4.3.1. Estimador del punto de cambio

Además de la condición (B1) consideramos las siguientes hipótesis.

(B0) La función b es continua y $b(T) = 0$.

(P1) La función $\mathrm{pen} : [0,T] \to \mathbb{R}_{\geq 0}$ es estrictamente creciente.

(P2) Existe $C_p > 0$ tal que si $\mathrm{pen}(\theta_2) - \mathrm{pen}(\theta_1) \geq C_p(\theta_2 - \theta_1)$ para todo $\theta_2 > \theta_1$.
Además $\mathrm{pen}(0) > 0$.

(B2) Existen $C_b > 0$ y $q > 0$ tales que

$$\alpha(\delta) := \inf_{\theta < \theta_0 - \delta} P\,\langle b - b_\theta\,,\, x - Px \rangle^2 \geq C_b \delta^q.$$

Si bien la penalización $\mathrm{pen}(\theta) = \theta^2$ no cumple la condición (P2) si lo cumple si se considera $\mathrm{pen}(\theta) = (\theta + c)^2$ para cualquier $c > 0$. Notemos también que la condición (B2) implíca la condición (B1).

Mediante el siguiente teorema y la subsiguiente proposición obtenemos tasas para la estimación del punto de cambio. No hemos encontrado trabajos que las obtengan en la literatura. Además, para la misma metodología de estimación de Hall y Hooker (2016) recuperamos su resultado de consistencia pero con un enfoque diferente en la demostración el cual creemos aporta una visión de la validez del resultado que puede ser aplicada en otros contextos.

Teorema 4.2 *Sean $b_n^\star$ una sucesión de estimadores de b y $\mathtt{a}_n$ una secuencia de números reales no negativa tales que*

$$S_n := \sup_{\theta \in [0,T]} |\ell(\theta, b_n^\star, P_n) - \ell(\theta, b, P)| = \mathcal{O}_P(\mathtt{a}_n), \tag{73}$$

Si $\omega_n = o(1)$, $\mathtt{a}_n = o(\omega_n)$ y además valen las condiciones (B0), (B1) y (P1) entonces $\hat{\theta}_n$ converge a θ_0 en probabilidad.

Más aún, si se satisfacen adicionalmente las condiciones (P2) y (B2), tomando $\omega_n = \mathtt{a}_n^{q/(q+1)}$ resulta que

$$|\hat{\theta}_n - \theta_0| = \mathcal{O}_P\left(\mathtt{a}_n^{1/(q+1)}\right). \tag{74}$$

Dem: 10.1

En lo que sigue, mostramos bajo que condiciones se satisface la hipótesis 73 del Teorema 4.2.

Proposición 4.3 *Si $b_n^\star$ es una sucesión de estimadores de b y satisface $\|b_n^\star\|^2 = \mathcal{O}_P(1)$ y $\sup_{\theta\in[0,T]} \|b_\theta^\star - b_\theta\|_{\Gamma_n} = \mathcal{O}_P(\mathtt{b}_n)$ entonces*

$$\sup_{\theta\in[0,T]} |\ell(\theta, b_n^\star, P_n) - \ell(\theta_0, b, P)| = \mathcal{O}_P(\text{máx}\{\mathtt{b}_n,\, n^{-1/2}\}). \tag{75}$$

Dem: 10.2

La siguiente observación nos muestra como obtener las tasas de convergencia de la estimación del punto de cambio si conocemos las tasas de convergencia en norma L_2 del estimador inicial. Estas tasas las conocemos para el estimador sin restricciones sobre un espacio de splines gracias a la Proposición 3.24.

Observación 4.4 *Si se satisface la condición (A2) por Cauchy-Schwartz tenemos*

$$\sup_{\theta\in[0,T]} \|b_\theta^\star - b_\theta\|_{\Gamma_n}^2 \leq \frac{4}{n}\sum_{i=1}^{n} \|x_i - \overline{x}\|^2 \sup_{\theta\in[0,T]} \|b_\theta^\star - b_\theta\|^2 = \mathcal{O}_P(\|b^\star - b\|^2).$$

Observación 4.5 *No es posible controlar el término $\sup_{\theta\in[0,T]} \|b_\theta^\star - b_\theta\|_\Gamma$ a partir de $\|b^\star - b\|_\Gamma$ sin imponer condiciones poco naturales sobre el operador de covarianza del proceso aleatorio.*

4.3.2. Estimación en tres pasos

En esta subsección estudiamos las tasa de convergencia en Γ-Norma para los estimadores a tres pasos $\widehat{b}_n[\hat{\theta}_n]$ y $\widetilde{b}_n[\hat{\theta}_n]$. Hasta nuestro conocimiento ni sus tasas ni su consistencia han sido estudiadas. Para esto hacemos uso de las herramientas desarrolladas en la primera parte de la tesis, en particular el Teorema 2.14 y la Observación 2.16. Estos nos dicen que para obtener las tasas en Γ-Norma solo necesitamos obtener tasas en norma empírica y mostrar que las normas L_2 de las estimaciones están acotadas en probabilidad.

Nuestro primer resultado muestra que si $\widehat{b}$ es definido por un subespacio H entonces, independientemente del tamaño de la muestra, la suma de errores cuadrados penalizados del estimador truncado cerca de θ_0 depende del espacio H sólo a través de su dimensión y su capacidad para aproximar a la función de peso truncada. Necesitamos la siguiente condición sobre el proceso que genera las covariables.

(A5) Las covariables tienen derivada continua con probabilidad uno y además el proceso $w(t) = t(x'(t) - m'(t))$ es un $L_2[0,T]$ elemento aleatorio.

Está condición se cumple si por ejemplo $\mathtt{E}\left(\|x'\|^2 < +\infty\right)$.

Teorema 4.6 *Sea $(x_1, Y_1), \ldots (x_n, Y_n)$ una muestra aleatoria del modelo 64 que satisface (A5). Sea $H \subseteq L_2[0,1]$ un subespacio de dimensión finita mediante el cual se define $\widehat{b}[\theta]$ por la Ecuación 71. Si $0 < \delta < \theta_0/4$ entonces existen variables aleatorias no negativas B_n e $Y_{H,\delta,\rho,n}$ tales que*

$$\sup_{\theta \in [\theta_0 - \delta, \theta_0 + \delta]} \|\widehat{b}_n[\theta] - b\|^2_{\Gamma_n} + \rho \|\widehat{b}_n[\theta]\|^2 \leq Y_{H,\delta,\rho,n} + \frac{\delta}{\rho} B_n,$$

donde B_n no depende de H, ρ ni δ y

$$B_n = \mathcal{O}_P(1).$$

Además, si p denota la dimensión del espacio H tomando

$$C = \text{máx}\{\sigma^2, \|b\|^2, 4\mathtt{E}(\|x - m\|^2)(1 + \|b\|^2)\},$$

el cuál no depende de H, ρ, δ ni n, vale

$$\mathtt{E}(Y_{H,\delta,\rho,n}) \leq C \left(\frac{p}{n} + \sup_{\theta \in [\theta_0 - \delta, \theta_0 + \delta]} \|\Pi_{H_\theta}(b_\theta) - b_\theta\|^2 + \rho + \delta \right).$$

Dem: 10.3

Para los siguientes resultados asumimos que $\hat{\theta}_n$ es una sucesión de estimadores de θ_0 y que $\mathtt{d}_n$ es una sucesión de números reales positivos tales que

$$|\hat{\theta}_n - \theta_0| = \mathcal{O}_P(\mathtt{d}_n). \tag{76}$$

Como $\hat{\theta}_n$ depende de las covariables no hacemos ningún tipo de supuesto de independencia entre estas. La siguiente proposición establece las tasas de convergencia siempre que la sucesión de espacios H_n tengan buenas propiedades de aproximación.

Proposición 4.7 *Sean $\mathtt{d}_n$ definida en la Ecuación 76, $H_n \subseteq L_2[0,1]$ una sucesión de subespacios de dimensión p_n tales que*

$$\text{máx}\left\{ \sup_{\theta \in [0,T]} \|\Pi_{H_{n\theta}}(b_\theta) - b_\theta\|^2, \, \frac{p_n}{n} \right\} = \mathcal{O}\left(\mathtt{d}_n^{1/2}\right).$$

Si se satisface la condición (A5) y ρ_n es una sucesión de parámetros suavizados tales que

$$\rho_n \propto \mathtt{d}_n^{1/2},$$

entonces

$$\|\widehat{b}_n[\hat{\theta}_n] - b\|^2_{\Gamma_n} = \mathcal{O}_P(\mathtt{d}_n^{1/2}) \qquad y \qquad \|\widehat{b}_n[\hat{\theta}_n]\|^2 = \mathcal{O}_P(1).$$

Dem: 10.4

El siguiente resultado muestra que podemos utilizar estimación restringida para el tercer paso si asumimos que b pertenece a $\mathcal{C} \subseteq L_2[0,T]$ un conjunto convexo de funciones continuas que se anulan a partir de θ_0. Este espacio podría ser funciones decrecientes hasta θ_0 o también cóncavas en el intervalo $[0, \theta_0]$.

Corolario 4.8 *Sean $H_n \subseteq L_2[0,1]$ subespacios de dimensión p_n y $E_{n\theta} = \overline{\mathcal{C} \cap H_{n\theta}}$. Si $b \in \mathcal{C}$ y ρ_n es una sucesión de parámetros de suavizado que satisface*

$$\text{máx}\left\{ \sup_{\theta \in [0,T]} \|\Pi_{H_{n\theta}}(b_\theta) - b_\theta\|^2, \frac{p_n}{n} \right\} = \mathcal{O}\left(\mathtt{d}_n^{1/2}\right).$$

Si se satisface la condición (A5) y ρ_n es una sucesión de parámetros suavizados tales que

$$\rho_n \propto \mathtt{d}_n^{1/2},$$

entonces

$$\|\widetilde{b}_n[\hat{\theta}_n] - b\|^2_{\Gamma_n} = \mathcal{O}_P(\rho_n) \qquad y \qquad \|\widetilde{b}_n[\hat{\theta}_n]\|^2 = \mathcal{O}_P(1).$$

Dem: 10.5

Finalmente la siguiente observación nos da los elementos necesarios para trabajar con espacios de splines y nos muestra que los estimadores propuestos en esta tesis son consistentes para la estimación de tres pasos y que podemos obtener sus tasas.

Observación 4.9 *Si $b \in C^k[0,T]$ resulta que $b|_{[0,\theta]} \in C^k[0,\theta]$. Si $H = S_p^k[0,1]$ es el espacio de splines con grilla uniforme del $[0,1]$ y dimensión p entonces se satisface que $\tilde{H}_\theta = S_p^k[0,\theta]$. Por el Teorema 3.22 si $k = 3$ o es par tenemos*

$$\|\Pi_{H_\theta}(b_\theta) - b_\theta\|^2 \leq \theta \| \, b|_{[0,\theta]} - \pi_{\tilde{H}_\theta} \, b|_{[0,\theta]} \, \|^2_\infty \leq 2D_k \frac{T^{2k+1}}{(p-k)^{2k}} \propto p^{-2k}.$$

Por lo tanto tomando $p_n = n^{1/(2k+1)}$ obtenemos

$$\text{máx}\left\{ \sup_{\theta \in [0,T]} \|\Pi_{H_{n\theta}}(b_\theta) - b_\theta\|^2, \frac{p_n}{n} \right\} = \mathcal{O}\left(n^{-2k/(2k+1)}\right).$$

5. Estudio de simulación

En esta sección presentamos un estudio de simulación realizado con el fin de estudiar el comportamiento en muestras finitas de los modelos y estimadores presentados en los Capítulos 3 y 4. Este se divide en dos partes. La primera la dedicamos a evaluar el comportamiento de los estimadores del Modelo Lineal con Restricciones y la segunda a los del Modelo Lineal Histórico.

Para generar muestras (x_i, Y_i) con $1 \leq i \leq n$ que sigan un modelo lineal

$$Y_i = \alpha + \int_0^1 b(t)x(t)\,dt + \varepsilon_i,$$

necesitamos definir una ley de distribución de las covariables x, una función de regresión b, un intercept α y una distribución para los errores ε. En todos los escenarios de la simulación utilizamos $\alpha = 0$ y errores $\varepsilon \sim \mathcal{N}(0, \sigma^2)$.

1. Proceso de Fourier q.

   Los elementos del proceso se obtienen como una combinación lineal infinita

   $$a_0 + \sum_{k=1}^{\infty} a_{2k-1}\sin(2k\pi t) + a_{2k}\cos(2k\pi t),$$

   donde los coeficientes $a_j \sim \mathcal{N}\left(0, (\frac{2}{j})^{2q}\right)$ si $j \geq 1$ y $a_0 \sim \mathcal{N}(0,1)$ son mutuamente independientes. En este proceso tenemos $\mu(t) = \mathtt{E}(x(t)) = 0$ y los autovalores λ del proceso serán los valores 1 y $(\frac{2}{j})^{2q}$ con $j \in \mathbb{N}$.

   Para generar una muestra elegimos un valor de truncado $M \in \mathbb{N}$ y generamos la normal multivariada correspondiente para obtener los a_k

   $$x(t) = a_0 + \sum_{k=1}^{M} a_{2k-1}\sin(2k\pi t) + a_{2k}\cos(2k\pi t).$$

   Para generar n elementos del proceso utilizaremos como valor M al entero mas cercando a máx$\{25, n \times 0{,}65\}$. Esta elección fue tomada para garantizar que el espacio donde viven los elementos del proceso es de dimensión mayor al generado por las n covariables.

2. Proceso Ornstein-Uhlenbeck.

   Generamos una muestra del proceso definido en 1.8 utilizando una grilla equiespaciada del dominio y de tamaño 100. Interpolamos con una poligonal continua los valores.

3. Proceso Punch.

   Los elementos de este proceso tienen la siguiente forma

   $$x(t) = Ze^{w^{-2}(t)} + X,$$

   donde Z y X son variables aleatorias independientes. La función $w(t)$ también es generada de manera aleatoria. Esta es una función lineal a trozos con un punto de quiebre cuya ubicación

se define sorteando de una variable U con soporte incluido en el intervalo $(0,1)$. La función queda determinada por un valor fijo en ambos extremos del dominio y cero en el punto de quiebre.

Este proceso presenta lo que se conoce en la literatura como variación de fase. Esto es, la variación del proceso se presenta en el eje del tiempo. En este caso el tiempo donde ocurre el máximo de la función. Incluimos este escenario ya que muchas de las metodologías de datos funcionales suelen presentar problema ante este tipo de variación.

Consideramos las siguientes funciones de regresión con dominio el $[0,1]$ re-escaladas para que tengan norma 1.

Coseno Truncado:	$b(t) = (\cos(2\pi t) - 1)\mathbb{I}_{[0,0,5]}(t)$;
Escalón:	$b(t) = \mathbb{I}_{[0,0,49]}(t) + 100(0{,}5 - t)\mathbb{I}_{(0,49,0,5]}(t)$;
Cauchy:	$b(t) = (t - 0{,}5)^2 e^{-0{,}01(t-0{,}5)^{-2}}\mathbb{I}_{[0,0,5]}(t)$;
Homográfica:	$b(t) = (t + 0{,}1)^{-1} - (1{,}1)^{-1}$;
Nula:	$b(t) = 0$;
Polinomio Centrado:	$b(t) = (t - 0{,}5)^6 - (0{,}5)^6$;
Cuadrática:	$b(t) = (0{,}5 - t)^2\mathbb{I}_{[0,0,5]}(t)$;
Lineal:	$b(t) = (0{,}5 - t)\mathbb{I}_{[0,0,5]}(t)$;
Pico:	$b(t) = t\mathbb{I}[0, 0{,}2](t) + (1{,}2 - 3t)\mathbb{I}(0{,}2, 0{,}4](t)$.

Las Figura 7 muestra sus respectivos gráficos. Como características observables, todas las funciones valen 0 en 1 y son no negativas.

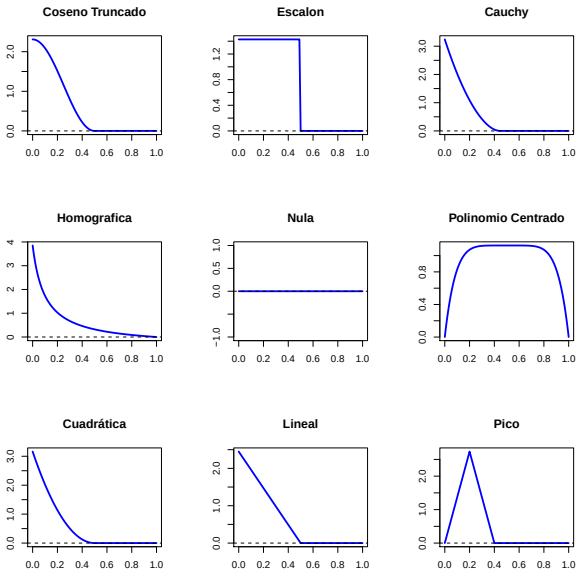

Figura 7: Funciones de regresión utilizadas en el estudio de simulación.

5.1. Estimación de la función de peso bajo restricciones de forma

En el presente estudio utilizamos cuatro métodos de generación de covariables. Dos procesos de Fourier con $q = 1$ y $q = 0{,}2$, un proceso Ornstein-Uhlenbeck con $\mu = 0$, $\theta = 1$ y $\sigma = 0{,}5$ y un proceso Punch con $Z \sim \mathcal{N}(0, 1/49)$, $X \sim \mathcal{N}(0, 0{,}01)$ y $U \sim \mathcal{U}(0{,}1, 0{,}9)$. En la Figura 8 se exhibe una muestra aleatoria de 6 elementos para cada uno de estos procesos. En rojo grueso se grafica la media del proceso obtenida al promediar 2000 elementos.

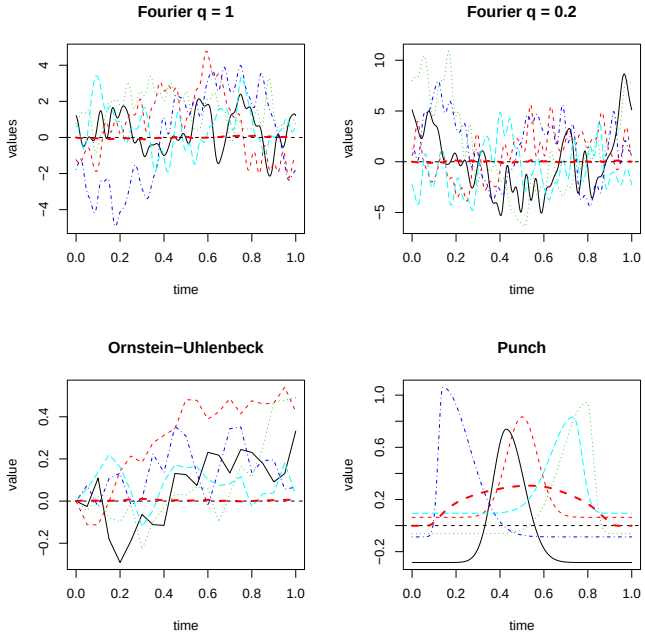

Figura 8: Muestra aleatoria de 6 elementos para cada proceso de generación de covariables. En rojo y trazo grueso se grafica la media del proceso.

Cada escenario de simulación se corresponde a una de las 4 metodologías y una función de regresión. Dentro de cada escenario se generaron muestras para $n \in \{5, 10, 15, 20, 25, 50, 100, 200\}$. Se utilizó $\sigma^2 = 1$ para la varianza de los errores.

Dependiendo la función de regresión se usaron distintas restricciones. Con las funciones Cauchy y Homográfica restricciones de Monotonía y de Convexidad. Con la función Nula restricciones de Monotonía y Concavidad. Para las funciones Escalón y Coseno Truncado se utilizaron restricciones de Monotonía y con las funciones Polinomio Centrado y Pico restricciones de Concavidad. Las funciones Lineal y Cuadrática se utilizan para el Modelo Lineal Histórico.

Para cada tipo de restricción existen diferentes variantes: Monótona y Monótona0, Convexa y Convexa0, Cóncava, Cóncava0 y Cóncava00. Más precisamente, la restricción que llamamos Monótona es de decrecimiento. Cuando figura un 0 en la restricción se impone que el valor de la función estimada debe ser 0 en el borde derecho. La presencia de 00 corresponde a valer 0 en ambos extremos del dominio. Esta última se utiliza solamente con las funciones Nula, Polinomio Centrado y Pico.

Como espacios aproximantes H consideramos los espacios S_p^3 y S_p^4. Como se desarrolla en la Subsec-

ción 3.5.1, el espacio de splines de orden 3 y dimensión p lo utilizamos para estimar bajo restricciones de Monotonía. El espacio de splines de orden 4 y dimensión p lo utilizamos para obtener restricciones de Concavidad o Convexidad. Para los valores de p consideramos dos variantes. Una fija que no depende del tamaño de la muestra y otro que depende del n. Estos son $p = 10$ y $p = \lfloor n/2 \rfloor$

Con el objetivo de comparar y apreciar los resultados consideramos dos tipos de estimadores sin restricciones. El estimador Pca basado en componentes principales funcionales y un estimador Irrestricto. Este último se calcula mediante la Ecuación 43 usando como espacio aproximante el mismo del estimador restringido con el que se compara.

También tuvimos en consideración dos tipos de penalizaciones. La penalización Ridge $\text{Pen}_{RI}(f) = \|f\|^2$ y una penalización a la derivada segunda $Pen_{SD}(f) = \|f^{(2)}\|^2 + c_p\|\Pi_{\text{lineales}}(b)\|^2$. El primer términco de la penalización favorece soluciones suaves mientras que el último término es para que la penalización resulte inyectiva. El valor c_p corresponde al autovalor mínimo del operador derivada segunda dentro de S_p^k. El operador Π_{lineales} es el proyector sobre el núcleo de la derivada segunda. Es decir el proyector sobre el espacio de funciones lineales.

Por cada tipo de restricción, dimensión del espacio aproximante y penalidad, tenemos un estimador diferente. Teniendo en cuenta que los estimadores Irrestricto con S_p^3 e Irrestricto con S_p^4 son distintos contamos un total de $9 \times 2 \times 2 + 1 = 37$ estimadores.

Para elegir el parámetro de suavizado ρ de cada estimador y la cantidad de componentes principales utilizada por el estimador Pca se realiza una validación cruzada de 5 cruces. Para la elección de ρ utilizamos una grilla lo suficientemente amplia. La cantidad de componentes principales para el estimador Pca se limita entre 1 y el mínimo entre 15 y la dimensión del espacio generado por la muestra.

Para cuantificar el error en cada estimación utilizamos la norma cuadrada y la Γ-norma cuadrada de $\widehat{b} - b$. Esta última la estimamos generando una muestra de tamaño $B = 1200$ y calculamos

$$\|\widehat{b} - b\|^2_{\Gamma_B} = \frac{1}{B}\sum_{i=1}^{B} \left\langle \widehat{b} - b\,,\, x_i - \bar{x} \right\rangle^2 .$$

Reportamos el promedio de estos valores en $N = 1200$ replicaciones de la simulación. Obtenemos así estimaciones de $\mathtt{E}(\|\widehat{b} - b\|^2)$ y $\mathtt{E}(\|\widehat{b} - b\|^2_\Gamma)$. Esta última estimación queda justificada por la Ley de los Grandes Números Uniformes si B es grande y $\mathtt{E}(\|\widehat{b}\|^2) < \infty$ resulta

$$\mathtt{E}(\|\widehat{b} - b\|^2_\Gamma) \approx \mathtt{E}(\|\widehat{b} - b\|^2_{\Gamma_B}).$$

Para resumir los resultados obtenidos se presentan gráficos y tablas que se presentan en el Apéndice 11.2. Los gráficos los presentamos en escala log-log. Una característica importante de este tipo de gráfico es que las funciones de la forma $f(n) = cn^\alpha$ se grafican como rectas de ecuación $y = \alpha x + \log(c)$.

Las Figuras 18 a 33 y la Figura 50 muestran los gráficos en escala log-log de los errores de estimación medidos para $\mathtt{E}\left(\|\widehat{b}_n - b)\|^2_\Gamma\right)$. Las Figuras 20 a 34 muestran los gráficos en escala log-log de la simulación cuantificados por $\mathtt{E}\left(\|\widehat{b}_n - b\|^2\right)$.

Las Tablas 1 a 9 muestran las diferencias en los resultados de estimar con restricciones de monotonía, concavidad o convexidad. Cada uno corresponde con un proceso de generación de covariables y a una norma con la que se cuantifican los errores.

Cada fila corresponde a una estimador (salvo las filas recuadradas). Cada columna corresponde a una cantidad de covariables. En el interior de la tabla figuran la proporción de aumento del error respecto a la metodología que mejor resultado obtuvo para ese columna. Dicho de otro modo, si llamamos $\widehat{b}_n^i$ al estimador correspondiente a la fila i para un tamaño de muestra n en la entrada (i, n) de la tabla tenemos

$$\frac{\|\widehat{b}_n^i - b\|_J^2}{\text{mín}_j \|\widehat{b}_n^j - b\|_J^2} - 1,$$

donde J puede ser 2 o Γ. En la fila recuadrada anotamos en la columna que corresponde a n el valor $\text{mín}_j \|\widehat{b}_n^j - b\|_J^2$.

5.1.1. Algunas observaciones sobre los errores cuantificados por la Γ-Norma

Las siguientes conclusiones se aplican sobre todos los escenarios de simulación salvo los que incluyen a la función Pico. Las observaciones se desprenden de las Figuras 18 a 33.

- Como es de esperar, las curvas de las estimaciones son decrecientes. Es decir, el error decrece cuando el n aumenta.
- Si comparamos las curvas de igual p e igual penalización observamos que en todos los escenarios obtenemos, como es de esperar, que a mayor nivel de restricción mejor resulta la estimación. Por ejemplo, las curvas de error de predicción de Monótona0 mejoran uniformemente a las de Monótona que a su vez mejoran uniformemente a las de Irrestricto.
- En casi todos los escenario, para todo $n \geq 10$, las estimaciones obtenidas con cualquier restricción bajo la penalización Ridge, mejoran uniformemente a las estimaciones por Pca. Las únicas excepciones corresponden a algunos escenarios con la función nula y solo para $n = 10$.
- A partir de $n = 10$ las curvas se asemejan a rectas.
- Los resultados de los procesos Fourier con $q = 1$ y $q = 0{,}2$ son muy similares para los escenarios comparables. Las Tablas 2 y 3 permiten apreciar esto con más detalle para las funciones Nula, Cauchy y Homográfica.
- En ambos procesos Fourier, para todas las funciones de regresión salvo la Nula, las curvas de los estimadores restringidos se ven poco afectadas por la elección de la penalización. Los mejores resultados por penalización dependen en gran medida de la función de regresión.
- En los 4 casos de generación de covariables la estimación de la función Nula se ve beneficiada por el uso de penalización Ridge. Es decir, fija la restricción y el p, los errores con penalización Ridge son uniformemente mejores a los de la estimación con penalización Segunda Derivada. Esto es de esperar por la naturaleza de la penalización.

- En los escenarios con Ornstein-Uhlenbeck y Punch bajo restricciones de monotonía, las curvas de error son uniformemente mejores para la penalización Ridge sobre la penalización Segunda Derivada para $n \leq 100$. Se observa que para todas las funciones salvo la Nula pareciera existir una tendencia en las curvas de error de Monótona0 con penalización Segunda Derivada que intersecan o eventualmente intersecarían a la de la penalización Ridge. Esta tendencia no se observa para las estimaciones de Irrestricto.
- El escenario con la función Nula y covariables Fourier $q = 1$ y $q = 0{,}2$ el error para el estimador Pca es casi constante para $10 \leq n \leq 25$. En el caso $q = 0{,}2$ para $n = 5$ el error del estimador Pca es menor que con $n = 10$.
- Salvo para la función Nula las curvas son similares entre cada función a estimar. El proceso que genera las covariables parece tener mayor influencia sobre las curvas de error que las funciones de regresión.
- Para la función Nula las curvas de error tienen forma similar entre los distintos procesos que generan las covariables.

5.1.2. Algunas observaciones sobre los errores cuantificados por la norma L_2

Las siguientes observaciones se desprenden de las Figuras 34 a 49.

- En ambos procesos de Fourier la estimación en norma es estrictamente decreciente para los estimadores con espacio de dimensión fija $p = 10$.

  Si comparamos las curvas para cada método con distintas restricciones ($p = 10$ e igual penalización) observamos, en todos los escenarios, que a mayor nivel de restricción mejor es la estimación.
- En ambos procesos de Fourier el comportamiento de los estimadores con $p = n/2$ es diverso. Si bien en algunos escenarios es decreciente y se acerca a 0 en otros no resulta monótono ni parece acercarse a 0 a medida que aumenta el n.
- Los procesos Ornstein-Uhlenbeck y Punch el error de estimación medido en norma está muy lejos de 0 para todas las funciones y restricciones. El comportamiento de algunos estimadores es errático.
- Los procesos Fourier presentan resultados similares para $q = 1$ y $q = 0{,}2$. Esto también se puede apreciar en las Tablas 6 y 7.
- Las curvas de Irrestricto y Monótona con $p = n/2$ y penalización Segunda Derivada resultan estrictamente crecientes en el proceso Punch. Esto sugiere que

  $$\sup_{n \in \mathbb{N}} \mathtt{E}\left(\|\widehat{b}_n - b\|^2 \right) = +\infty.$$

  Notemos que la penalización Segunda Derivada no es coerciva como sucesión de penalizaciones sobre S_p^3. Esto puede incidir en los valores de la esperanza de la norma del estimador como se desprende del Teorema 3.10 y luego del Teorema 3.13.

De todas formas para que $\mathtt{E}\left(\|\widehat{b}_{n_0} - b\|^2_\Gamma\right) \approx \frac{1}{B}\sum_{i=1}^{B}\left\langle \widehat{b}_{n_0} - b\,,\, x_i - \bar{x}\right\rangle^2$ siga valiendo solo necesitamos que para cada n_0 se satisfaga $\mathtt{E}\left(\|\widehat{b}_{n_0} - b\|^2\right) < +\infty$. Es interesante el contraste con la Figura 25 donde los errores medidos en Γ-Norma decrecen a 0 a medida que el n crece.

- Para todos los $n \geq 50$ y casi todos los $n \geq 20$ los estimadores restringidos con $p = 10$ mejoran al estimador Pca.
- El estimador Pca no parece acercarse a 0 a medida que aumenta la cantidad de muestras.

5.1.3. Estimación con restricciones cuando no se cumplen las condiciones

La Figura 50 muestra los resultados de estimar bajo restricciones erróneas en el Modelo Lineal Funcional.

La función Pico no es cóncava y podemos observar que las restricciones Concava0 y Concava00 dan resultados muy pobres en comparación con el estimador Pca e Irrestricto. A pesar de que la metodología Concava tiene menos información correcta sobre la función (no conoce el valor de la función de regresión en ninguno de los bordes del domino) es la de menor error. La estimaciones empeoran sustantivamente en comparación a las estimaciones sin restricciones al aumentar el tamaño de la muestra. Llamativamente la estimación Concava da los mejores resultado para tamaños de muestra menores a 25 en el proceso de Fourier $q = 1$.

5.1.4. Estimación Monótona vs estimación Convexa

De la información proporcionada por las Tablas 2 a 9 observamos que para el error medido en Γ-Norma y en norma suele ser mejor utilizar información de la monotonía que de la convexidad. Esto puede atribuirse a que se tiene más información de una función a partir de su derivada primera que de su derivada segunda.

5.2. Estimación en el Modelo Lineal Histórico

La segunda parte de la simulación corresponde estudiar el comportamiento de los distintos estimadores propuestos para el Modelo Lineal Histórico. Utilizamos dos de los procesos de generación de covariables ya descriptos. Un proceso de Fourier $q = 1$ y un proceso Ornstein-Uhlenbeck con $\mu = 0$, $\theta = 1$ y $\sigma = 0{,}5$. Para la intercept y la varianza de los errores del modelo utilizamos $\alpha = 0$ y $\sigma^2 = 0{,}7$.

Para funciones de regresión usamos las funciones Cauchy, Cuadrática, Lineal y Escalón. Todas estas tienen punto de cambio $\theta_0 = 0{,}5$ y son decrecientes. Además están ordenadas de mayor a menor en cantidad de derivadas continuas en el punto de truncado. Esto corresponde con el valor q de la hipótesis (B2), siendo $q = \infty$, 2, 1 y 1 respectivamente.

Consideramos dos estimadores iniciales, el estimador Pca y el estimador Monótona0 con penalidad Ridge y $p = 10$. La determinación de la cantidad de componentes principales y del parámetro de suavizado ρ se eligen mediante una validación cruzada de 5 cruces con toda la muestra.

Como función de penalización para el punto de truncado utilizamos $\text{pen}(\theta) = \theta^2$. El criterio de elección fue su facilidad para implementar el algoritmo computacional. Para estimar θ_0 a partir del estimador inicial realizamos validación cruzada de 5 cruces para elegir el parámetro de suavizado ω. Evaluamos su desempeño estimando el error de predicción del estimador inicial truncado en $\hat{\theta}(\omega)$.

Luego de estimar el punto de corte calculamos un estimador de 3 pasos. Para su cálculo estimamos con restricción Monótona0 con $p = 10$ y penalidad Ridge sobre el intervalo $[0, \hat{\theta}]$. Luego extendemos la función para que esté definida en todo el intervalo $[0, 1]$. Medimos el desempeño de estos estimadores con la Γ-Norma cuadrada y la norma cuadrada.

Las Tablas 10 y 11 muestran las medidas de resumen para las distribuciones de los errores de estimación del punto de corte. Las Figuras 51 a 58 muestran histogramas que comparan la distribución de los errores discriminados por función de regresión y cantidad de covariables. La Figura 59 presenta histogramas para comparar el desempeño en términos de la Γ-Norma del estimador inicial Monótona y su estimador a 3 pasos correspondiente.

La Tabla 12 compara los errores de estimación medidos en Γ-Norma para los dos estimadores de 3 pasos, los dos estimadores iniciales y el estimador Oráculo que es un estimador que conoce el valor verdadero del punto de cambio θ_0. Este estimador no es aplicable en la práctica, pero sirve como elemento de comparación sobre cuanto lugar queda para la mejora de los estimadores presentados.

En las filas recuadradas figura el error de estimación medido por Γ-Norma para el estimador a 3 pasos con inicial Monótona. En el resto de las filas corresponden a un estimador y las columnas a una cantidad de covariables. En el interior de la tabla figura la diferencia relativa entre el estimador a 3 pasos con inicial Monótona contra el estimador que corresponde a la fila y la cantidad de covariables de la columna. Es decir, si llamamos $\widehat{b}_n^i$ al estimador i para el tamaño de muestra n tenemos en cada entrada de la tabla

$$\frac{\|\widehat{b}_n^i - b\|_\Gamma^2}{\|\widehat{b}_n^{\text{mon 3 pasos}} - b\|_\Gamma^2} - 1.$$

Valores negativos en la tabla quieren decir que el estimador en cuestión obtuvo un mejor resultado que el estimador a 3 pasos con inicial Monotona. Es de esperar que el estimador Oráculo mejore a todos los estimadores ya que posee más información sobre la función a estimar.

5.2.1. Estimación del punto de corte

Las siguientes observaciones se desprenden de las Tablas 10 y 11 además de las Figuras 51 a 58.

- Como es de esperar, en todos los escenarios con covariables Fourier el error cuadrático medio decrece con el tamaño de muestra.
- En todos los escenarios con covariables Fourier la estimación del punto de corte con el estimador inicial Monotona son uniformemente mejores a las estimaciones por Pca en términos del error cuadrático medio.
- El error cuadrático medio del estimador mejora a medida que disminuye la cantidad de derivadas continuas de la función de peso en el punto u_0. Esto corresponde con la hipótesis (B2) utilizada para obtener las tasas de convergencia.

- En los escenarios con covariables Fourier $q = 1$ la estimación del punto de corte con el estimador inicial Monótona pareciera ser sesgado.
- Con el estimador inicial Monotona y covariables Fourier, la variabilidad de los errores disminuye con la cantidad de derivadas continuas en u_0.
- En los escenarios con covariables Ornstein-Uhlenbeck presenta mejores resultados para el estimador inicial por Pca. Sin embargo los histogramas muestran para ambos iniciales que un intento de estimar u_0 resulta fútil. En efecto, El primer cuartil de todos los errores es $-0{,}5$ que es el error máximo que se puede tener. En la estimación del punto de corte con inicial Monotona tenemos que en la mayoría de los escenarios incluso el segundo cuartil es cercano a $-0{,}5$.

  Una posible justificación para el mal desempeño de los estimadores lleva a considerar la cantidad $\sup_{\theta\in[0,T]} \|(\widehat{b} - b)\mathbb{I}_{[0,\theta]}\|_{\Gamma_n}$. Garantizar que este término este controlado es una de las hipótesis de la Proposición 4.3. Como se desprende de la Observación 4.4 este término se puede acotar por la norma del error de estimación. De las Figuras 34 y 38 vemos que la norma de los errores para los estimadores iniciales Pca y Monótona son mucho más chicas en el proceso Fourier que en los procesos Ornstein-Uhlenbeck.

5.2.2. Estimación en tres pasos de la función de regresión

Las siguientes obsrvaciones sobre los estimadores de tres pasos se desprenden de la Tabla 12 y la Figura 59.

- El error de estimación medido en Γ-norma para el estimador Pca siempre se ve beneficiado por una estimación de 3 pasos. Esto es de alguna manera esperable porque en el estimador inicial Monotona es uniformemente mejor.
- Con covariables Fourier la estimación de 3 pasos y la estimación a un paso con inicial Monótona presentan resultados similares sin terminar de favorecer a uno sobre el otro. Las diferencias relativas son menores al 10 porciento. Por lo tanto una estimación de 3 pasos podría ser interesante por su interpretabilidad.
- Con covariables Ornstein-Uhlenbeck se ve que una estimación a 3 pasos mejora el error. Sin embargo hemos visto que la estimación del u_0 no es fiable por este método.

5.3. Conclusiones

De la presente simulación concluimos que siempre que exista información sobre la monotonía o la concavidad de la función de regresión es conveniente utilizar procedimientos de estimación que contemplen estas propiedades. También apreciamos su beneficio para la estimación del punto de cambio en el Modelo Lineal Histórico cuando las covariables siguen un proceso de Fourier. Entre las opciones de penalidades estudiadas la Ridge parece ser más apropiada para los tamaños de muestra moderados aunque en varios escenarios la penalización Segunda Derivada resultó uniformemente mejor.

6. Ejemplo de aplicación: Emisión de partículas sólidas

En esta sección, con el objetivo de ilustrar las propuestas de estimación de esta tesis, estudiamos los datos de emisión de partículas sólidas (PS) de camiones a combustible diesel. Estos datos fueron presentados en Clark (2007) y se utilizaron para ilustrar el Modelo Lineal Histórico en Hall (2014) y Guan (2020).

Las datos se obtuvieron colocando un contador de partículas en el caño de escape de camiones y midiendo a cada segundo t la cantidad de partículas sólidas emitidas $PS(t)$ y la velocidad del camión a lo largo de un ciclo de conducción. En Ascencio (2014) se propuso el siguiente modelo para los datos

$$\log(PS(t)) = \int_0^{\theta_0} b(u)Z(t-u)\,du,$$

donde $Z(t)$ es la aceleración del motor en el segundo t. Esto indica que la cantidad de partículas emitidas en un tiempo t depende sólo de los θ_0 segundos previos a su emisión. El parámetro θ_0 es de interés y por argumentos técnicos es menor o igual a 60. La respuestas Y se obtuvieron para distintos valores de t suficientemente espaciados para disminuir dependencias entre las curvas y otras respuestas.

Disponemos de $n = 108$ curvas de velocidad W_i evaluadas en una grilla del intervalo $[0, 60]$ sobre los números enteros y la cantidad de partículas sólidas contabilizadas para cada una. Construimos las covariables $X_i(t) = -W_i'(60 - t)$ asociadas a sus respuestas Y_i. Es decir $X_i(0)$ es la aceleración del motor en el momento que se contabilizaron las partículas sólidas y $X(60)$ la aceleración 60 segundos antes.

En la Figura 9 presentamos un histograma de las variables Y_i ya centradas y las curvas de aceleración también centradas. El esquema de colores utilizado pinta a cada curva del color que corresponde a su respuesta Y_i en una paleta que va de Rojo a Amarillo para valores de menor a mayor. Por la cantidad de datos es difícil obtener información del gráfico de la totalidad delas curvas, aunque se puede apreciar zonas de color amarillo para valores altos de aceleración en los primeros 20 segundos. En las Figura 10 presentamos multiples gráficos cada uno con 4 curvas seleccionadas al azar manteniendo el esquema de colores. La Figura 10 presenta los gráficos de algunas curvas seleccionadas al azar y discriminadas por el valor de respuesta. De estos gráficos podemos observar que las curvas con $Y > 10$, es decir, para las que más partículas sólidas se les contabilizaron, presentan valores positivos de aceleración entre 0 y 20. Pareciera que valores altos de emisión de partículas se asocian a aceleraciones positivas en los 20 segundos anteriores. En las curvas con $Y < -10$ toman valores positivos entre los 20 y 40 segundos antes de la medición de la cantidad de partículas.

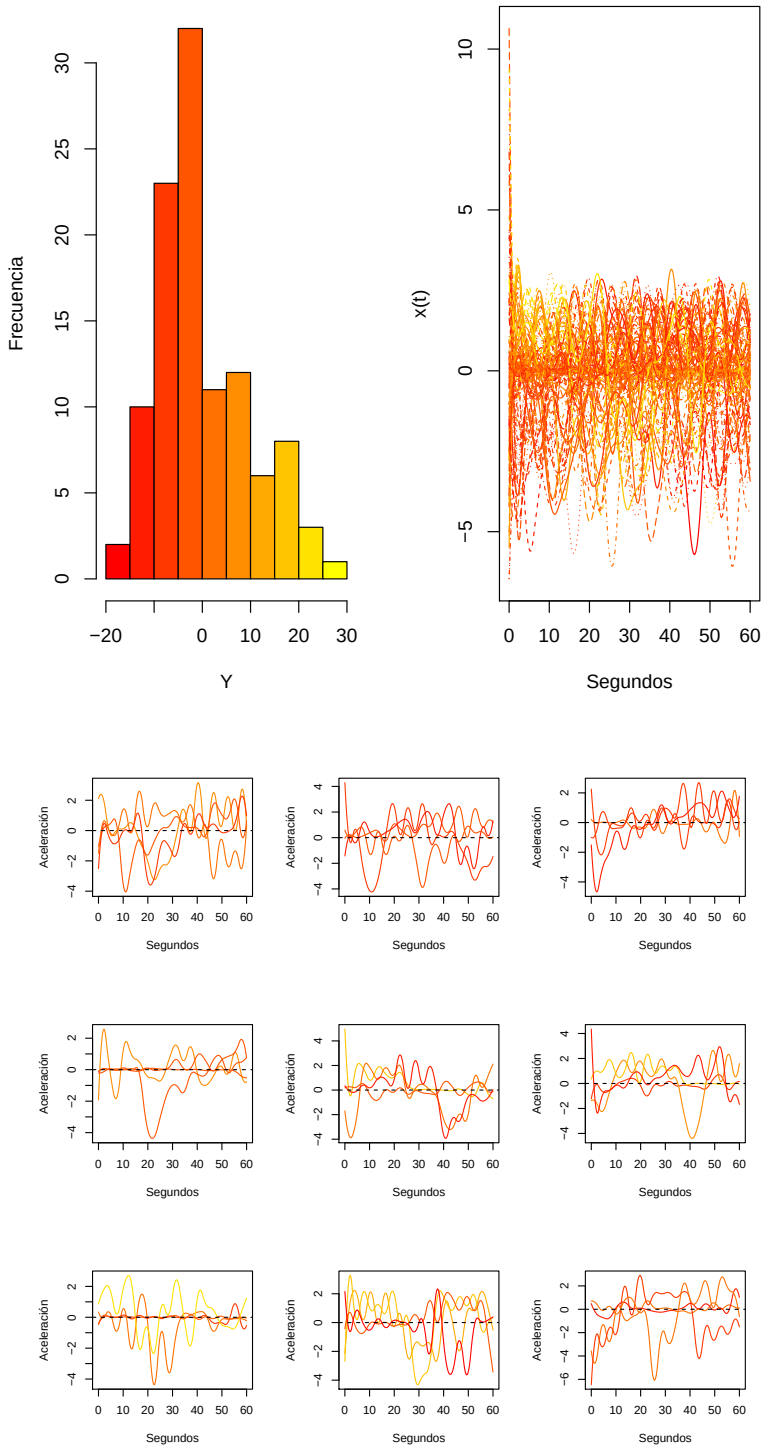

Figura 9: Histograma de los valores de Y centrados y las curvas de aceleración centradas con su color correspondiente al de su respuesta en el histograma.

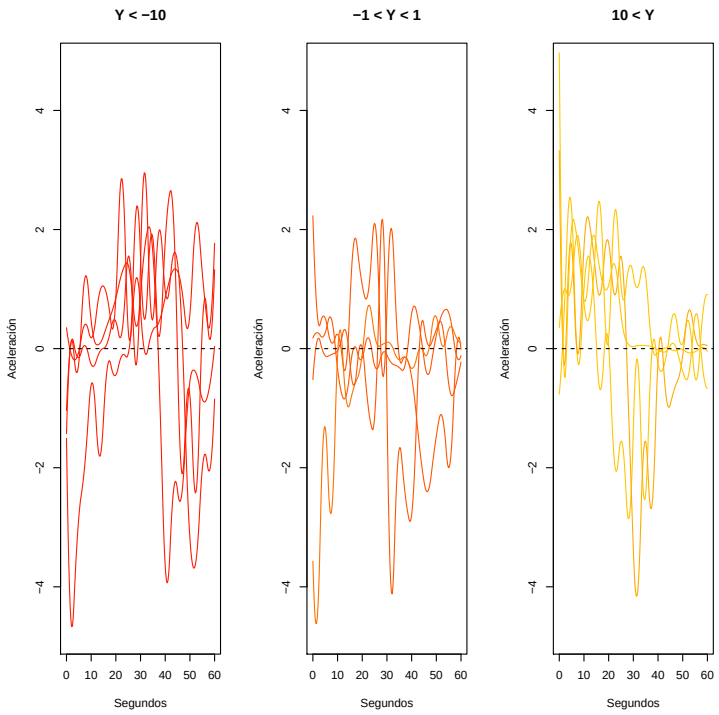

Figura 10: Algunas curvas de aceleración discriminadas por el valor de su respuesta.

Para estudiar los datos ajustamos el modelo lineal con restricciones Decreciente0 e Irrestricto0. El modelo decreciente podría tener sentido si uno supone que la influencia de la aceleración en la cantidad de partículas sólidas disminuye a medida que aumenta el tiempo transcurrido desde la medición. Con esta restricción la función estimada será decreciente y nula en el borde derecho. La restricción Irrestricto0 solo asume que la función de peso borde derecho de su dominio es nula. Utilizamos las penalizaciones Ridge y Segunda Derivada obteniendo 4 estimaciones distintas. También ajustamos el modelo mediante estimación por componentes principales principales funcionales. Los parámetros de suavizado y la cantidad de componentes principales las elegimos mediante validación cruzada de 5 cruces. Las funciones de peso estimadas se presentan en las Figura 11 y las curvas de la validación cruzada en las Figuras 14 y 14. De los estimaciones Irrestricto0 y Pca se ve reflejado lo observado en el párrafo anterior. Valores positivos de aceleración de 0 a 20 contribuirían a mayores valores de partículas sólidas y valores positivos de 20 a 40 contribuirían a menor cantidad de partículas sólidas. Esto no se ve reflejado por el estimador Decreciente0 que se anula cerca del valor 20 y según este solo influiría en la cantidad de partículas sería los valores de aceleración de 0 a 20.

Utilizando estos estimadores como iniciales aplicamos validación cruzada de 5 cruces para determinar el parámetro de suavizado en la estimación del punto de truncado en el modelo lineal histórico. Los valores estimados se presentan en la Tabla 1. Para la validación cruzada dividimos los datos en cinco grupos. En cada cruce utilizamos 4 de ellos para determinar el punto de truncado correspondiente a cada parámetro de suavizado. Para cada punto de corte ajustamos el modelo lineal con Irrestricto0 sobre el dominio restringido y evaluamos su desempeño en el grupo restante. Utilizamos la misma penalización con la cual se calculó el estimador inicial correspondiente y utilizamos como

	Dec0	Irr0	Pca
Ridge	20.4	39.6	39.0
SD	14.2	40.2	39.0

Tabla 1: Estimación del punto de truncado para cada estimador inicial y penalización utilizada en la re-estimación del modelo.

parámetro de suavizado el que se obtuvo por validación cruzada para el inicial ajustado para tener en cuenta el dominio más pequeño. En el caso del estimador inicial por componentes principales reajustamos con ambas penalizaciones. Las curvas de error de validación cruzada se presentan en las Figura 15.

De la validación cruzada pareciera ser que Irrestrico0 y Pca son más adecuados como estimadores iniciales ya que obtienen un error de reconstrucción estimado menor que Decreciente0.

La Figura 12 muestra las estimaciones a tres pasos. Se puede apreciar que se mantienen las formas de las estimaciones presentadas en la Figura 11. Esto podría indicar que un punto de truncado alrededor de 40 es razonable. Este valor es considerablemente diferente a los obtenidos en Hall y Hooker (2016) y Guan et al. (2020) donde estiman el punto de truncado al rededor de 14 y 20 respectivamente. Estos valores son los que obtuvimos con la restricción Decreciente0 bajo las distintas penalidades.

En la Figura 16 se presentan las relaciones entre el parámetro de suavizado, el punto de truncado y el error de reconstrucción relativo. La primera columna de gráficos muestra el punto de truncado estimado como función del parámetro de suavizado ω, es decir

$$\theta(\omega) = \underset{u\in[0,60]}{\operatorname{argmin}} \sum_{i=1}^{168} \left(Y_i - \int_0^u \widehat{b}(t) x_i(t)\, dt \right)^2 + \omega u^2.$$

Si definimos el error de reconstrucción para cada ω

$$E(\omega) = \sum_{i=1}^{168} \left(Y_i - \int_0^{\theta_0(\omega)} \widehat{b}(t) x_i(t)\, dt \right)^2.$$

los gráficos de la segunda columna presentan el error relativo de reconstrucción, es decir, $E(\omega)/E(0)$. De manera similar definimos el error relativo para cada posible punto de truncado

$$Er(u) = \sum_{i=1}^{168} \left(Y_i - \int_0^u \widehat{b}(t) x_i(t)\, dt \right)^2.$$

la tercer columna presenta $Er(u)/Er(60)$. Los colores de cada curva se mantienen como en la Figura 11 para representar a cada estimador inicial. En los gráficos de la primera columna se observan saltos verticales en las curvas. Esto se debe a que el punto de truncado $\theta(\omega)$ no es una función continua del parámetro de suavizado. Para entender que es lo que sucede presentamos la Figura 17 donde se grafican las curvas de error de reconstrucción penalizado $u \to Er(u) + \omega u^2$ para distintos valores del parámetro de suavizado ω. De estos se puede apreciar que las discontinuidades se deben a la existencia de máximos locales o cambios abruptos de concavidad en la curva con $\omega = 0$.

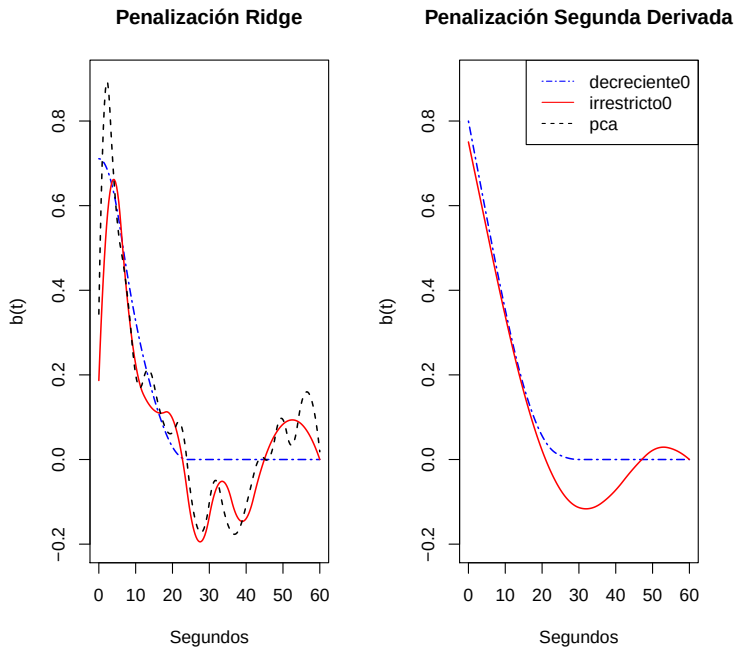

Figura 11: Estimación de funciones de regresión.

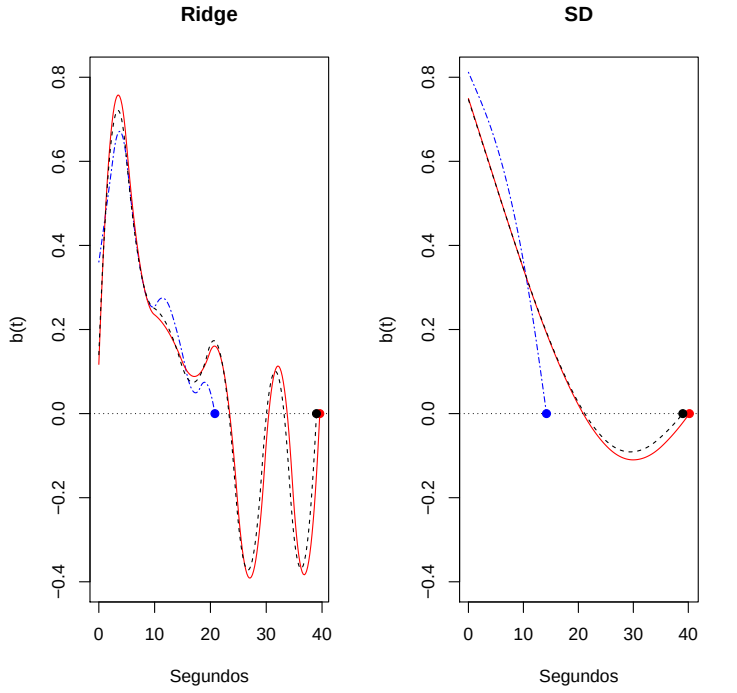

Figura 12: Estimación de la función de regresión en tres pasos.

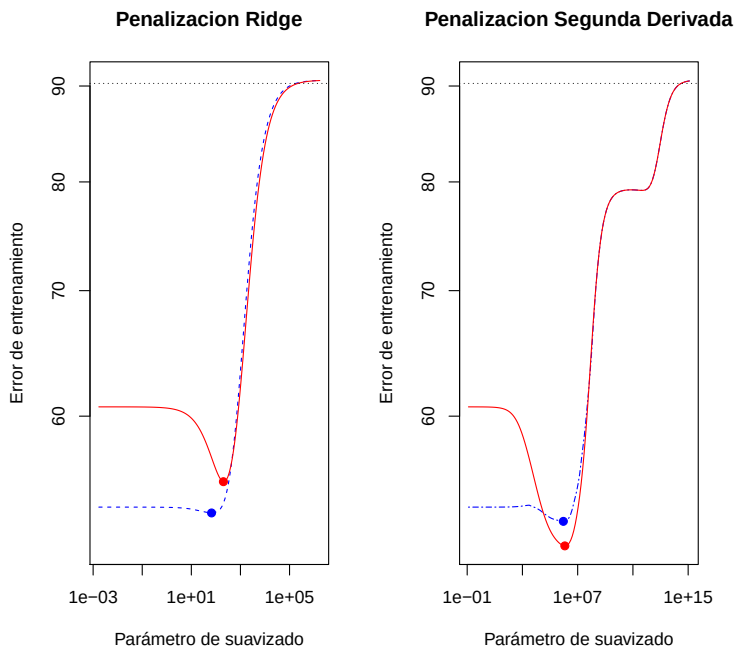

Figura 13: Validación cruzada para la estimación de las funciones de regresión.

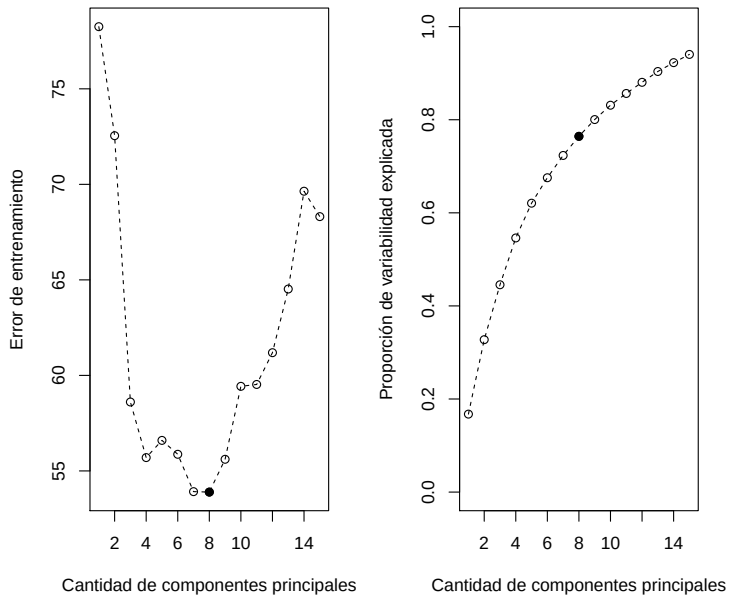

Figura 14: Validación cruzada y variabilidad explicada para el estimador Pca.

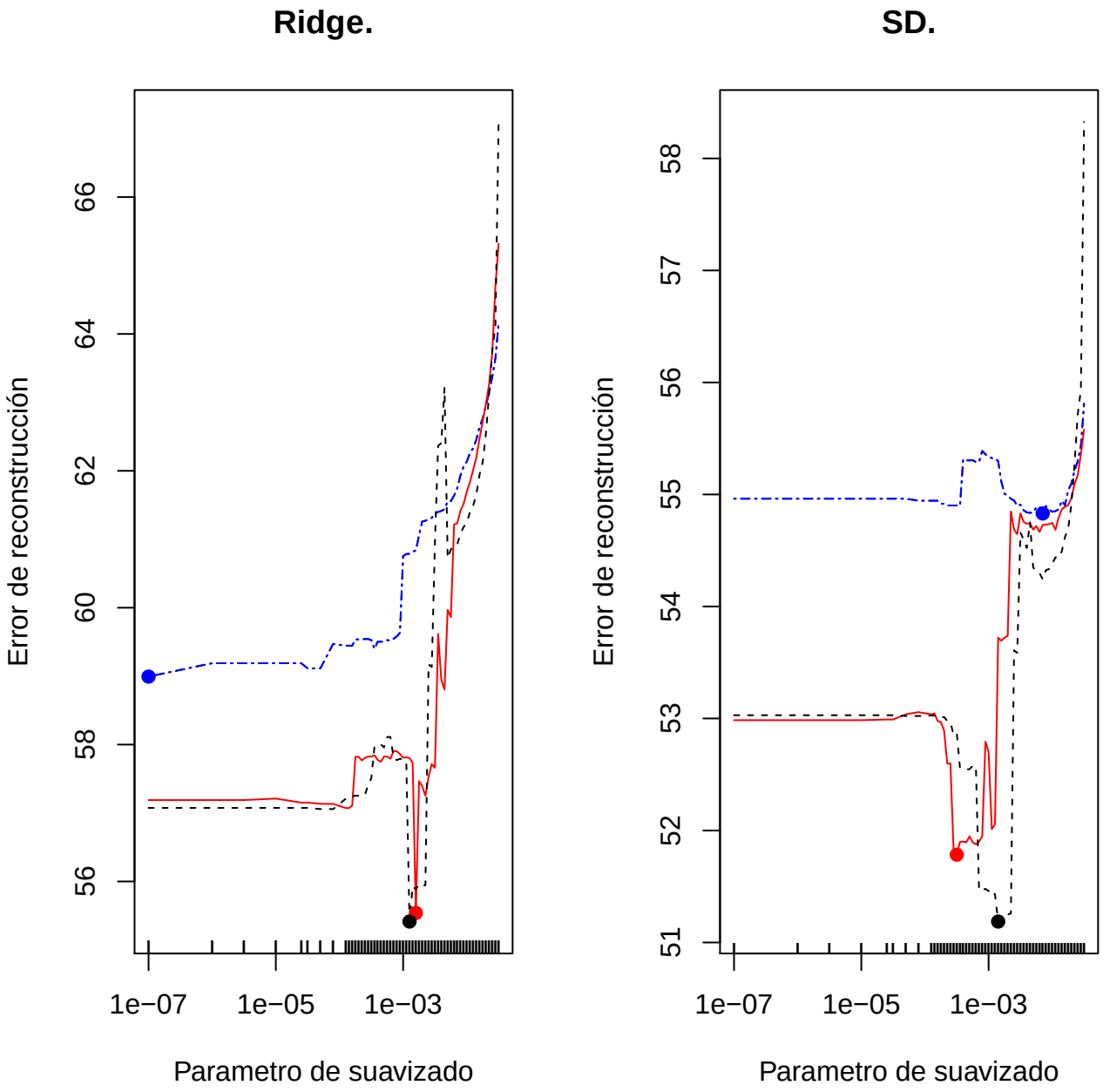

Figura 15: Curva de validación cruzada para estimaciones iniciales y componentes principales funcionales. Los puntos indican donde se alcanza el mínimo de cada curva.

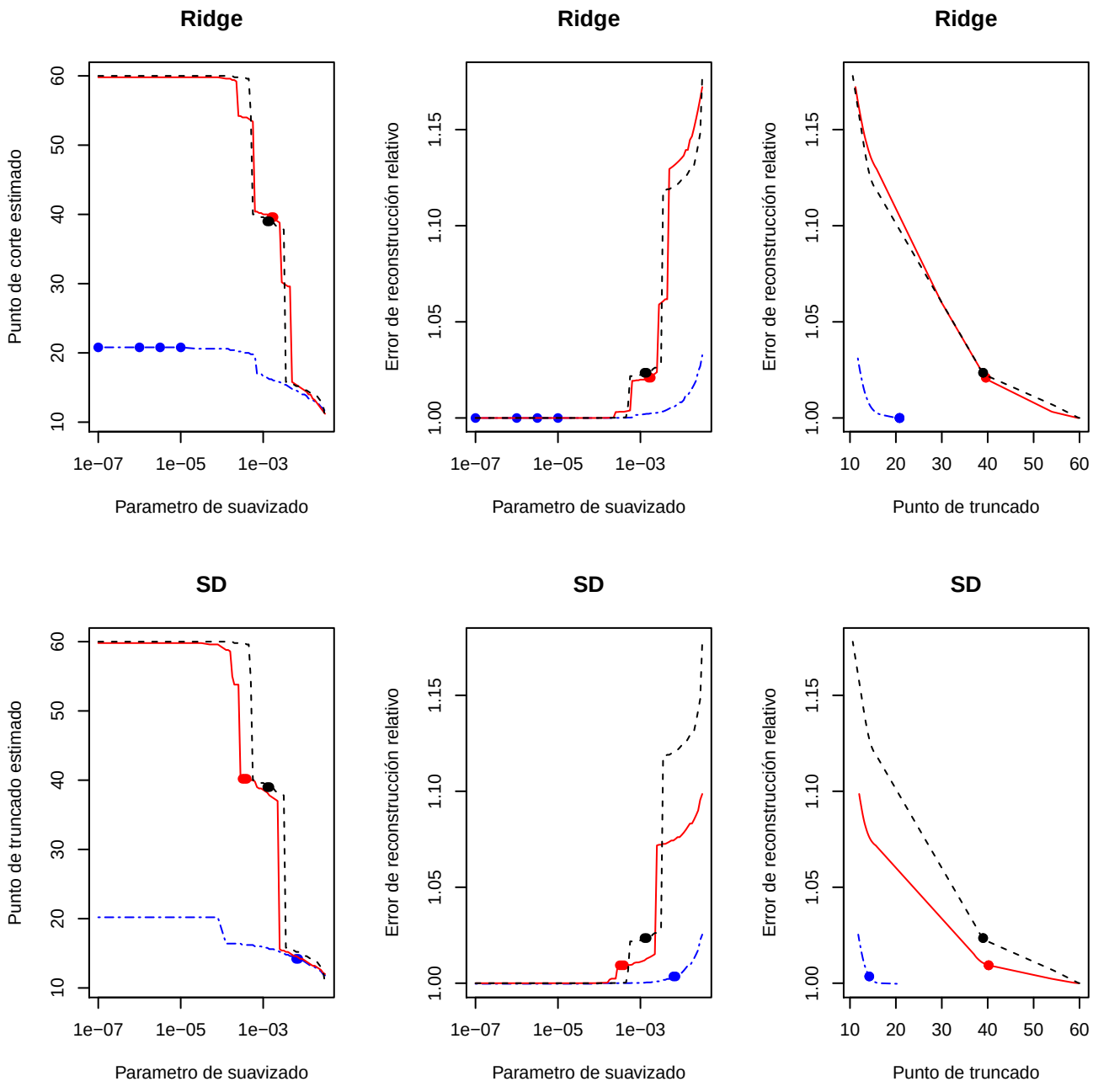

Figura 16: Relaciones entre el punto de truncado estimado el parámetro de suavizado y error de reconstrucción relativo para estimaciones iniciales con la penalización Ridge y por componentes principales funcionales. Los puntos corresponden a los puntos de truncado que se obtienen por la validación cruzada y/o a los parámetros de suavizado con los que se obtienen.

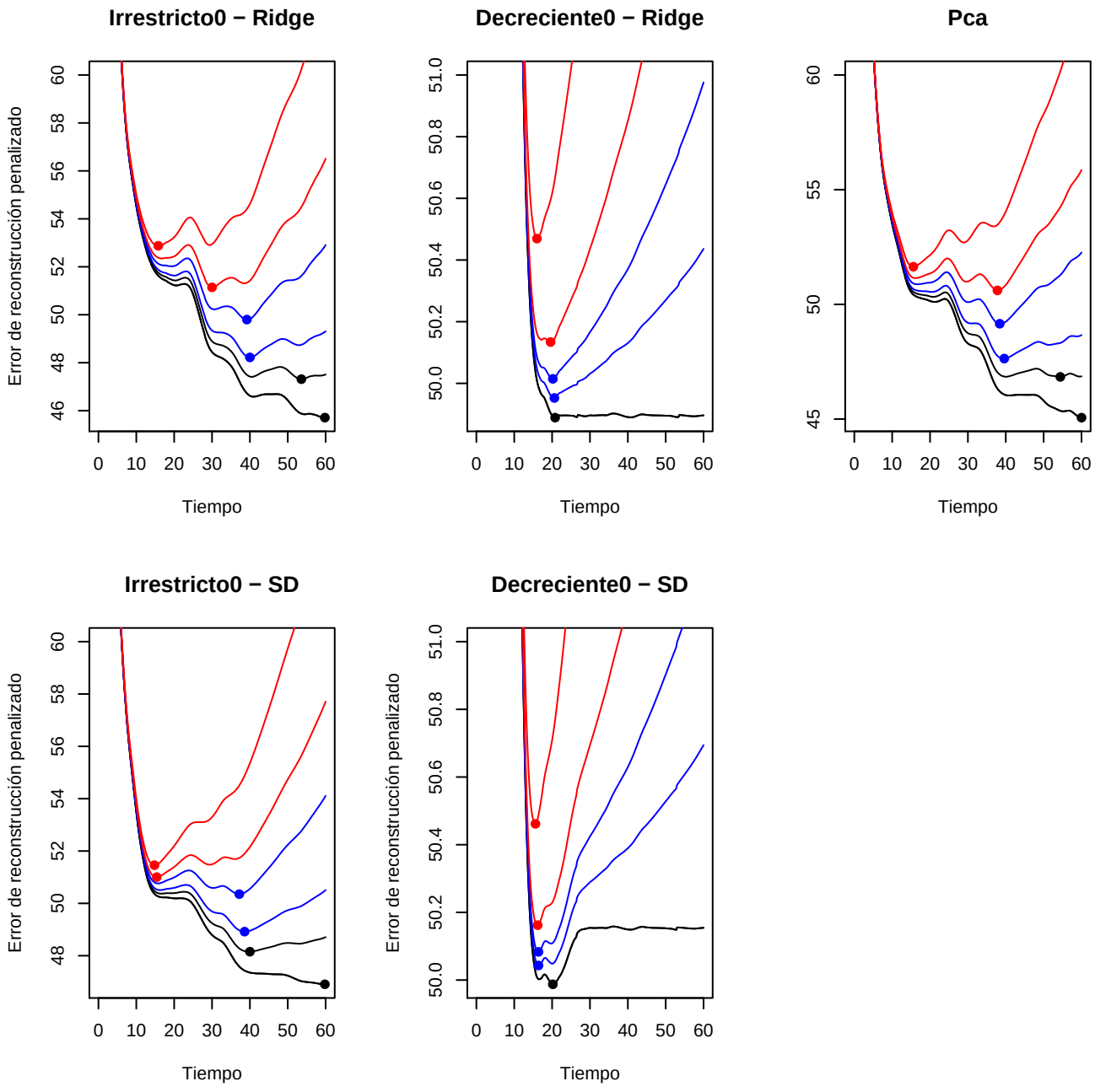

Figura 17: Error de reconstrucción penalizado en función del punto de truncado. Cada curva corresponde a un valor del parámetro de suavizado $\omega = 0, 0{,}0005, 0{,}001, 0{,}002, 0{,}003, 0{,}005$ salvo los gráficos Decreciente0 con $\omega = 0, 0{,}00015, 0{,}0003, 0{,}0006, 0{,}0018$. Los puntos sobre cada curva corresponden a su valor mínimo. Los colores no indican nada solo están para una mejor visualización.

7. Demostraciones del Capítulo 1

7.1 (Teorema 1.24) *Podemos asumir $m = 0$ sin perdida de generalidad. Notemos que $\chi_n = \langle x, e_n\rangle$ es una variable aleatoria en $\mathbb{R}$ por definición de un $\mathbb{H}$ elemento aleatorio. Por otro lado $\mathtt{E}(\langle x, e_n\rangle^2) \leq \mathtt{E}(\|x\|^2\|e_n\|^2) < \infty$. De lo que concluimos que $|\langle x, e_n\rangle| < +\infty$ con probabilidad 1. Claramente $\mathtt{E}(\chi_n) = 0$, usando el Teorema 1.23 obtenemos que $\mathtt{E}(x\chi_n) = \mathtt{E}(x\langle x, e_n\rangle) = \Gamma(e_n) = \lambda_n e_n$ y por lo tanto*

$$\begin{aligned} E(\chi_n\chi_k) &= \mathtt{E}(\langle x, e_n\rangle\langle x, e_k\rangle), \\ &= \langle \mathtt{E}(\langle x, e_n\rangle x), e_k\rangle, \\ &= \lambda_n\delta_{n,k}. \end{aligned}$$

Por último, utilizando los resultados obtenidos, tenemos la velocidad de convergencia de la serie truncada en término de los autovalores del operador de covarianza.

$$\begin{aligned} \mathtt{E}(\|x - \sum_{n=1}^{N}\chi_n e_n\|^2) &= \mathtt{E}(\|\sum_{n=N+1}^{\infty}\chi_n e_n\|^2) \\ &= \mathtt{E}(\sum_{n=N+1}^{\infty}\chi_n^2) \\ &= \sum_{n=N+1}^{\infty}\lambda_n. \end{aligned}$$

7.2 (Lema 1.25) *Como $\mathtt{E}\left(\int_I |x(t)u(t)|\,dt\right) \leq \mathtt{E}(\|x\|_2)\|u\|_2 < \infty$ por el Teorema de Fubini-Tonelli se puede intercambiar la esperanza con la integral obteniendo*

$$\mathtt{E}\left(\int_I x(t)u(t)\,dt\right) = \int_I \mathtt{E}(x(t))u(t)\,dt\,.$$

7.3 (Lema 1.26) *Podemos asumir $m = 0$ sin pérdida de generalidad. Como x es un $L_2(I)$ elemento aleatorio*

$$\begin{aligned} \Gamma[u](s) &= \mathtt{E}(\langle x, u\rangle x(s)), \\ &= \mathtt{E}\left(\int_a^b x(s)x(t)u(t)dt\right), \\ &= \int_a^b \mathtt{E}(x(s)x(t))u(t)\,dt\,. \end{aligned}$$

En la última igualdad se intercambia la esperanza con la integral por el Teorema de Fubini-Tonelli ya que $\mathtt{E}(\int_I |x(s)x(t)u(t)|\,dt) \leq \mathtt{E}(\|x\|_\infty\|x\|_2)\|u\|_2 \leq \mathtt{E}(\|x\|_2^2)^{1/2}\mathtt{E}(\|x\|_\infty^2)^{1/2}\|u\| < \infty$.

Para la segunda igualdad utilizamos el teorema de representación para un $L_2[a,b]$ elemento aleatorio

$$\begin{aligned}\mathtt{E}(x(s)x(t)) &= \mathtt{E}\left(\sum_{j=1}^{\infty}\chi_j e_j(s)\sum_{i=1}^{\infty}\chi_i e_i(t)\right), \\ &= \mathtt{E}\left(\sum_{j=1}^{\infty}\sum_{i=1}^{\infty}\chi_j\chi_i e_j(s)e_i(t)\right) = \sum_{j=1}^{\infty}\lambda_j e_j(s)e_j(t).\end{aligned}$$

8. Demostraciones del Capítulo 2

8.1 (Proposición 2.3) *Supongamos que no es cierto. Sea $v \in Nu(T)^{\perp}$ tal que $\langle u_n - u\,,\, v\rangle \not\to 0$ y sean $\varepsilon > 0$ y n_k tales que $|\langle u_{n_k} - u, v\rangle| > \varepsilon$. Como $\|u_{n_k} - u\|$ están acotadas, existen $w \in \mathbb{H}$ y n_{k_j} tales que $u_{n_{k_j}} - u \rightharpoonup w$. Como convergencia débil implica convergencia bajo T resulta que $u_{n_{k_j}} - u \xrightarrow{T} w$. Esto implica que $w \in Nu(T)$. Como $\langle w, v\rangle \geq \varepsilon$ llegamos a un absurdo.*

8.2 (Observación 2.7) *Claramente $\langle u, v\rangle_{\Gamma} = \langle \Gamma^{1/2}(u), \Gamma^{1/2}(v)\rangle$ es un producto interno y $\|u\|^2_{\Gamma} = \langle u, u\rangle_{\Gamma}$. Para ver que no es completo basta considerar $u_n = \sum_{k=1}^{n} e_k$. Esta sucesión es de Cauchy bajo la Γ-Norma. Si existiera límite $u \in \mathbb{H}$ debería satisfacer $\langle u, e_k\rangle = 1\, \forall k \in \mathbb{N}$.*

8.3 (Teorema 2.10) *Consideremos $\|u\|^2_{\Gamma_n^{\star}} = \frac{1}{n}\sum_{i=1}^{n}\langle u, x_i - m\rangle^2$, para cualquier $u \in \mathbb{H}$ resulta que*

$$\begin{aligned}
\|u\|^2_{\Gamma_n} &= \frac{1}{n}\sum_{i=1}^{n}\langle u, x_i - m + m - \bar{x}\rangle^2,\\
&= \frac{1}{n}\sum_{i=1}^{n}\langle u, x_i - m\rangle^2 + \frac{2}{n}\sum_{i=1}^{n}\langle u, x_i - m\rangle\langle u, m - \bar{x}\rangle + \frac{1}{n}\sum_{i=1}^{n}\langle u, m - \bar{x}\rangle^2,\\
&= \frac{1}{n}\sum_{i=1}^{n}\langle u, x_i - m\rangle^2 - 2\langle u, m - \bar{x}\rangle\langle u, m - \bar{x}\rangle + \langle u, m - \bar{x}\rangle^2,\\
&= \|u\|^2_{\Gamma_n^{\star}} - \langle u, m - \overline{x}\rangle^2.
\end{aligned}$$

Utilizando (A2) obtenemos que

$$\mathtt{E}\left(\sup_{\|u\|\leq 1}\left|\|u\|^2_{\Gamma} - \|u\|^2_{\Gamma_n^{\star}}\right|\right) = \mathtt{E}\left(\sup_{\|u\|\leq 1}\left|\langle u, m - \bar{x}\rangle^2\right|\right) = \mathcal{O}(\mathtt{m}_n). \tag{77}$$

Dado $u \in \mathbb{H}$ consideremos la secuencia $\mathtt{q}_n(u) = \max\left\{\|u\|_{\Gamma}\sqrt{\frac{\mathtt{t}_{k_n}}{n}}, \frac{\mathtt{s}_{k_n}}{\sqrt{n}}\right\}$. Afirmamos que vale la siguiente igualdad

$$\mathtt{E}\left(\sup_{\|u\|\leq 1} q_n(u)\left|\|u\|^2_{\Gamma_n^{\star}} - \|u\|^2_{\Gamma}\right|\right) = \mathcal{O}(1). \tag{78}$$

Para mostrar que esto es cierto, notemos que por el Teorema 1.24 $x = m + \sum_{r=1}^{\infty}\chi_r e_r$ y por lo tanto $\langle u, x - m\rangle^2 = \sum_{s=1}^{\infty}\sum_{r=1}^{\infty}\upsilon_r\upsilon_s\chi_r\chi_s$ con $\upsilon_r = \langle u, e_r\rangle$ obteniendo que

$$\begin{aligned}
\left|\|u\|^2_{\Gamma_n^{\star}} - \|u\|^2_{\Gamma}\right| &= \left|\frac{1}{n}\sum_{i=1}^{n}\sum_{r=1}^{\infty}\sum_{s=1}^{\infty}\upsilon_r\upsilon_s\chi_{ri}\chi_{si} - \sum_{r=1}^{\infty}\lambda_r\upsilon_r^2\right|,\\
&= \left|\sum_{r=1}^{\infty}\sum_{s=1}^{\infty}\upsilon_r\upsilon_s\left(\frac{1}{n}\sum_{i=1}^{n}\chi_{ri}\chi_{si} - \lambda_r I_{r=s}\right)\right|.
\end{aligned}$$

Definimos

$$\widetilde{\chi}_{rr} := \frac{1}{\lambda_r\sqrt{n}}\sum_{i=1}^{n}(\chi_{ri}^2 - \lambda_r) \qquad \widetilde{\chi}_{rs} := \frac{1}{\sqrt{\lambda_r\lambda_s n}}\sum_{i=1}^{n}\chi_{ri}\chi_{si},\ r \neq s \qquad \widetilde{\chi}_{rs} := 0 \ \textit{si}\ \min\{\lambda_r, \lambda_s\} = 0.$$

Tenemos que $\mathtt{E}(\widetilde{\chi}_{rs}) = 0$ *y si* $\min\{\lambda_r, \lambda_s\} > 0$ *por la hipótesis (A1) el segundo momento de estas variables está uniformemente acotado*

$$\mathtt{E}(\widetilde{\chi}_{rs}^2) = \frac{n}{\lambda_r\lambda_s}\mathtt{Var}\left(\frac{1}{n}\sum_{i=1}^{n}\chi_{ri}\chi_{si}\right) \leq C. \tag{79}$$

Continuamos acotando $|\|u\|^2_{\Gamma_n^\star} - \|u\|^2_\Gamma|$

$$\begin{aligned}
&\left|\sum_{r=1}^{\infty}\sum_{s=1}^{\infty}\upsilon_r\upsilon_s\left(\frac{1}{n}\sum_{i=1}^{n}\chi_{ri}\chi_{si} - \lambda_r I_{r=s}\right)\right| = \left|\frac{1}{\sqrt{n}}\sum_{r=1}^{\infty}\sum_{s=1}^{\infty}\upsilon_r\upsilon_s(\lambda_r\lambda_s)^{1/2}\widetilde{\chi}_{rs}\right| \\
&\leq \frac{2}{\sqrt{n}}\left|\sum_{r=1}^{k}\sum_{s=r}^{\infty}\upsilon_r\upsilon_s(\lambda_r\lambda_s)^{1/2}\widetilde{\chi}_{rs}\right| + \frac{2}{\sqrt{n}}\left|\sum_{r=k+1}^{\infty}\sum_{s=r}^{\infty}\upsilon_r\upsilon_s(\lambda_r\lambda_s)^{1/2}\widetilde{\chi}_{rs}\right| \\
&\leq \frac{2}{\sqrt{n}}\left(\sum_{r=1}^{k}\sum_{s=r}^{\infty}\lambda_r\upsilon_r^2\upsilon_s^2\right)^{1/2}\left(\sum_{r=1}^{k}\sum_{s=r}^{\infty}\lambda_s\widetilde{\chi}_{rs}^2\right)^{1/2} + \frac{2}{\sqrt{n}}\left(\sum_{r=k+1}^{\infty}\sum_{s=r}^{\infty}\upsilon_r^2\upsilon_s^2\right)^{1/2}\left(\sum_{r=k+1}^{\infty}\sum_{s=r}^{\infty}\lambda_s\lambda_r\widetilde{\chi}_{rs}^2\right)^{1/2}.
\end{aligned}$$

Tenemos que

$$\sum_{r=1}^{k}\sum_{s=r}^{\infty}\lambda_r\upsilon_r^2\upsilon_s^2 \leq \|u\|_\Gamma^2\|u\|^2 \quad y \quad \sum_{r=k+1}^{\infty}\sum_{s=r}^{\infty}\upsilon_r^2\upsilon_s^2 \leq \|u\|^4.$$

Usando la cota obtenida en 79 y Fubini-Tonelli resulta que

$$E\left(\sum_{r=1}^{k}\sum_{s=r}^{\infty}\lambda_s\widetilde{\chi}_{rs}^2\right) = \mathcal{O}(\mathtt{t}_k) \quad y \quad \mathtt{E}\left(\sum_{r=k+1}^{\infty}\sum_{s=r}^{\infty}\lambda_s\lambda_r\widetilde{\chi}_{rs}^2\right) = \mathcal{O}(\mathtt{s}_k^2). \tag{80}$$

Como la desigualdad se satisface para cualquier k podemos elegirlo en función del n siempre que $\frac{\mathtt{t}_{k_n}}{\sqrt{n}} \to 0$. *Utilizando la desigualdad de Jensen y que la raíz cuadrada es una función cóncava obtenemos 78.*

Para concluir la demostración sólo resta tomar esperanza y utilizar 77 y 78 en la siguiente desigualdad

$$\sup_{\|u\|\leq 1}\mathtt{f}_n^{-1}(u)\left|\|u\|_\Gamma^2 - \|u\|_{\Gamma_n}\right| \leq \sup_{\|u\|\leq 1}\mathtt{m}_n^{-1}\left|\|u\|_\Gamma^2 - \|u\|^2_{\Gamma_n^\star}\right| + \sup_{\|u\|\leq 1}\mathtt{q}_n^{-1}(u)\left|\|u\|^2_{\Gamma_n^\star} - \|u\|^2_{\Gamma_n}\right|.$$

8.4 (Corolario 2.12) *La demostración es análoga a la del Teorema 2.10. Basta notar que*

$$\mathtt{E}\left(\sup_{\|u\|\leq 1}\left|\|u\|_\Gamma^2 - \|u\|^2_{\Gamma_n^\star}\right|^2\right) = \mathtt{E}\left(\sup_{\|u\|\leq 1}\left|\langle u, m - \bar{x}\rangle^2\right|^2\right) = \mathcal{O}(\mathtt{l}_n^2),$$

y que

$$\left|\|u\|^2_{\Gamma_n^\star} - \|u\|^2_\Gamma\right|^2 \leq 8\|u\|^2 \sum_{r=1}^{k}\sum_{s=r}^{\infty}\lambda_s\widetilde{\chi}^2_{rs} + 8\|u\|^4 \sum_{r=k+1}^{\infty}\sum_{s=r}^{\infty}\lambda_s\lambda_r\widetilde{\chi}^2_{rs},$$

para luego utilizar las desigualdades obtenidas en la Ecuación 80.

8.5 (Observación 2.13) *Como Γ es nuclear por el Teorema 1.23 la sucesión $\mathtt{s}_k = \sum_{r=k+1}^{\infty}\lambda_r$ converge de manera decreciente a 0 mientras que t_k es creciente. Veamos que $\frac{\mathtt{t}_n}{n} = \frac{1}{n}\sum_{k=1}^{n}\mathtt{s}_k \to 0$. Notemos primero que $\frac{\mathtt{t}_n}{n}$ es una secuencia positiva y decreciente,*

$$\begin{aligned}\mathtt{t}_{n+1} - \mathtt{t}_n &= \frac{1}{n(n+1)}\left(n\sum_{k=1}^{n}\mathtt{s}_k - (n+1)\sum_{k=1}^{n}\mathtt{s}_k + n\mathtt{s}_{n+1}\right),\\ &= \frac{1}{n(n+1)}\left(n\mathtt{s}_{n+1} - \sum_{k=1}^{n}\mathtt{s}_k\right) < 0.\end{aligned}$$

Por lo tanto existe límite $L \geq 0$. Supongamos $L > 0$, existe un n_0 para el que $\mathtt{s}_n < L/2$ para todo $n \geq n_0$ pero

$$\frac{\mathtt{t}_n}{n} = \frac{1}{n}\sum_{k=1}^{n_0}\mathtt{s}_k + \frac{1}{n}\sum_{k=n_0+1}^{n}\mathtt{s}_k \leq \frac{n_0\mathtt{s}_1}{n} + \frac{n-n_0}{n}\frac{L}{2} \underset{n\to\infty}{\to} \frac{L}{2}.$$

Esto es una contradicción que vino de asumir $L > 0$.

8.6 (Teorema 2.14) *Primero notemos que*

$$\|u_n\|^2_\Gamma \leq \ |\|u_n\|^2_\Gamma - \|u_n\|^2_{\Gamma_n}| + \|u_n\|^2_{\Gamma_n}.$$

Llamamos $\widetilde{\mathtt{f}}_n(u) = \widetilde{\mathtt{f}}_n(\frac{u}{\|u\|})$ y obtenemos

$$\left|\|u_n\|^2_\Gamma - \|u_n\|^2_{\Gamma_n}\right| \leq \|u_n\|^2\widetilde{\mathtt{f}}_n(u_n)\sup_{\|u\|\leq 1}\frac{1}{\mathtt{f}_n(u)}\left|\|u\|^2_\Gamma - \|u\|^2_{\Gamma_n}\right|. \tag{81}$$

Hacemos uso del Teorema 2.10 y la propiedad $\mathcal{O}_P(a_n)\mathcal{O}_P(b_n) = \mathcal{O}_P(a_nb_n)$ para obtener

$$|\|u_n\|^2_{\Gamma_n} - \|u_n\|^2_\Gamma| = \mathcal{O}_P\left(\|u_n\|^2\widetilde{\mathtt{f}}_n\left(u_n\right)\right).$$

Como

$$\|u_n\|^2_\Gamma = \mathcal{O}_P\left(\text{máx}\left\{\mathtt{a}_n,\, \|u_n\|^2\widetilde{\mathtt{f}}_n\left(u_n\right)\right\}\right), \tag{82}$$

obtenemos finalmente

$$\|u_n\|^2\widetilde{\mathtt{f}}_n(u_n) = \text{máx}\left\{\sqrt{\frac{\mathtt{t}_{k_n}}{n}}\|u_n\|_\Gamma\|u_n\|,\, \frac{\mathtt{s}_{k_n}}{\sqrt{n}}\|u_n\|^2,\, \mathtt{m}_n\|u_n\|^2\right\}.$$

Resulta que

$$\begin{aligned}\|u_n\|^2_\Gamma &= \mathcal{O}_P\left(\text{máx}\left\{\mathtt{a}_n,\, \sqrt{\frac{\mathtt{t}_{k_n}}{n}}\|u_n\|_\Gamma\|u_n\|,\, \frac{\mathtt{s}_{k_n}}{\sqrt{n}}\|u_n\|^2,\, \mathtt{m}_n\|u_n\|^2\right\}\right),\\ &= \mathcal{O}_P\left(\text{máx}\left\{\mathtt{a}_n,\, \frac{\mathtt{t}_{k_n}}{n}\|u_n\|^2,\, \frac{\mathtt{s}_{k_n}}{\sqrt{n}}\|u_n\|^2,\, \mathtt{m}_n\|u_n\|^2\right\}\right).\end{aligned}$$

Para la última desigualdad usamos que $P(X \leq \max\{a,b\}) = P(X \leq a \cap X \leq b\}$ y por lo tanto al analizar la parte del máximo en que aparece $\|u_n\|_\Gamma$ podemos pasar dividiendo y luego elevar al cuadrado.

8.7 (Proposición 2.15) *Primero notemos que*

$$\|u_n\|_\Gamma \leq \ |\|u_n\|_\Gamma - \|u_n\|_{\Gamma_n}| + \|u_n\|_{\Gamma_n},$$

por lo tanto

$$|\|u_n\|_\Gamma - \|u_n\|_{\Gamma_n}| \leq \|u_n\| \sup_{\|u\|\leq 1} |\|u\|_\Gamma - \|u\|_{\Gamma_n}|.$$

Por la desigualdad triangular vale para todo $u \in \mathbb{H}$

$$|\|u\|_\Gamma - \|u\|_{\Gamma_n}|^2 \leq \left|\|u\|_\Gamma^2 - \|u\|_{\Gamma_n}^2\right|.$$

Usando Cauchy Schwartz, la última desigualdad y el Teorema 2.10

$$\begin{aligned}
\mathtt{E}(|\|u_n\|_\Gamma - \|u_n\|_{\Gamma_n}|) &\leq \mathtt{E}\left(\|u_n\|^2\right)^{1/2} \mathtt{E}\left(\sup_{\|u\|\leq 1} |\|u\|_\Gamma - \|u\|_{\Gamma_n}|^2\right)^{1/2}, \\
&\leq \mathtt{E}\left(\|u_n\|^2\right)^{1/2} \mathtt{E}\left(\sup_{\|u\|\leq 1} \left|\|u\|_\Gamma^2 - \|u\|_{\Gamma_n}^2\right|\right)^{1/2}, \\
&= \mathcal{O}(\max\left\{n^{-1/4}, \mathtt{m}_n^{1/2}\right\}).
\end{aligned}$$

De la misma manera

$$\|u_n\|_\Gamma^2 \leq \ |\|u_n\|_\Gamma^2 - \|u_n\|_{\Gamma_n}^2| + \|u_n\|_{\Gamma_n}^2,$$

y que

$$\left|\|u_n\|_\Gamma^2 - \|u_n\|_{\Gamma_n}^2\right| \leq \|u_n\|^2 \sup_{\|u\|\leq 1} \left|\|u\|_\Gamma^2 - \|u\|_{\Gamma_n}^2\right|,$$

podemos usar Cauchy Schwartz y la el Corolario 2.12

$$\begin{aligned}
\mathtt{E}(\left|\|u_n\|_\Gamma^2 - \|u_n\|_{\Gamma_n}^2\right|) &\leq \mathtt{E}\left(\|u_n\|^4\right)^{1/2} \mathtt{E}\left(\sup_{\|u\|\leq 1} \left|\|u\|_\Gamma^2 - \|u\|_{\Gamma_n}^2\right|^2\right), \\
&= \mathcal{O}\left(\max\left\{n^{-1/2}, \mathtt{l}_n\right\}\right).
\end{aligned}$$

8.8 (Ejemplo 2.17) *Para corroborar ambos ejemplos se debe acotar las series por las integrales análogas. En el item 2) considerar que $\mathcal{O}\left(\frac{k}{\ln^2(k)}\right) = \mathcal{O}\left(\frac{k}{\ln^2(k)} - \frac{2}{k\ln^3(k)}\right)$ si $k \geq 3$.*

8.9 (Proposición 2.18) *Este resultado se desprende del Teorema de Convergencia de Vitalli que dice que $\mathtt{E}(|X_n - X|) \to 0$ si y solo si $X_n \xrightarrow{p} X$ y X_n son uniformemente integrables.*

De la hipótesis $\sup_n \mathtt{E}(\|u_n\|_\Gamma^{2r}) < +\infty$ y la desigualdad

$$M\|u_n\|_\Gamma^r I_{\|u_n\|_\Gamma^r > M} \leq \|u_n\|_\Gamma^{2r},$$

obtenemos la uniforme integrabilidad de $\|u_n\|_\Gamma^r$ *ya que*

$$\sup_{n\in\mathbb{N}} \mathtt{E}(\|u_n\|_\Gamma I_{\|u_n\|_\Gamma^r > M}) \leq M^{-1} \sup_{n\in\mathbb{N}} \mathtt{E}(\|u_n\|_\Gamma^{2r}) \underset{M\to+\infty}{\longrightarrow} 0.$$

Para probar la segunda ecuación utilizamos con $r = 1$ *la Proposición 2.3 con* $T = \Gamma^{1/2}$*. En efecto*

$$\|\Gamma^{1/2}(\mathtt{E}(u_n))\| = \|\mathtt{E}\left(\Gamma^{1/2}(u_n)\right)\| \leq \mathtt{E}\left(\|\Gamma^{1/2}(u_n)\|\right) = \mathtt{E}\left(\|u_n\|_\Gamma\right),$$

El término $E(\|u_n\|_\Gamma)$ *ya probamos que tiende a cero. Además* $\|\mathtt{E}(u_n)\| \leq \mathtt{E}(\|u_n\|)$ *que es uniformemente acotado por hipótesis.*

8.10 (Proposición 2.19) *Por la Proposición 2.18 vale que* $\lim_{n\to+\infty} \mathtt{E}(\langle u_n - u_0\,,\, u_0\rangle) = 0$ *lo que implica* $\lim_{n\to+\infty} \mathtt{E}(\langle u_n\,,\, u_0\rangle) = \|u_0\|^2$*. Resulta entonces que*

$$\begin{aligned}
0 &\leq \liminf_n \mathtt{E}(\|u_n - u_0\|^2),\\
&\leq \limsup_n \mathtt{E}(\|u_n - u_0\|^2),\\
&= \limsup_n \mathtt{E}\left(\|u_n\|^2 + \|u_0\|^2 - 2\langle u_n\,,\, u_0\rangle\right),\\
&\leq 2\|u_0\|^2 - 2\lim_{n\to+\infty} \mathtt{E}(\langle u_n\,,\, u_0\rangle),\\
&= 0.
\end{aligned}$$

Como el límite superior e inferior coinciden y valen cero, concluimos que $\lim_{n\to+\infty} \mathtt{E}(\|u_n - u_0\|^2) = 0$*. La convergencia en* Γ*-Norma al cuadrado se desprende de este resultado.*

8.11 (Proposición) *2.20 Si* $u \in \mathcal{B}(\mathtt{v}, C)$ *resulta que para cualquier* $k \in \mathbb{N}$

$$\begin{aligned}
|\langle u_n\,,\, u\rangle| &\leq \sum_{j=1}^k |\langle u_n\,,\, e_j\rangle\langle u\,,\, e_j\rangle| + \sum_{j=k+1}^\infty |\langle u_n\,,\, e_j\rangle\langle u\,,\, e_j\rangle|,\\
&\leq \left(\sum_{j=1}^k \langle u_n\,,\, e_j\rangle^2\right)^{1/2}\left(\sum_{j=1}^k \langle u\,,\, e_j\rangle^2\right)^{1/2} + \left(\sum_{j=k+1}^\infty \langle u_n\,,\, e_j\rangle^2\right)^{1/2}\left(\sum_{j=k+1}^\infty \langle u\,,\, e_j\rangle^2\right)^{1/2}.
\end{aligned}$$

Como $\sum_{j=1}^k \langle u_n\,,\, e_j\rangle^2 \leq \lambda_k^{-1}\sum_{j=1}^k \lambda_j\langle u_n\,,\, e_j\rangle^2 \leq \lambda_k^{-1}\|u_n\|_\Gamma^2$ *obtenemos,*

$$\sup_{u\in\mathcal{B}(\mathtt{v},C)} |\langle u_n\,,\, u\rangle| \leq \|u_n\|_\Gamma \lambda_k^{-1/2} + C\|u_n\|\mathtt{w}_k.$$

Por lo tanto, utilizando las tasas y que la normas están acotadas en probabilidad o en primer momento obtenemos los resultados deseados.

8.12 (Corolario 2.21) *Tenemos que*

$$\begin{aligned}
\|u_n - u_0\|^2 &= \|u_n\|^2 + \|u_0\|^2 - 2\langle u_n\,,\, u_0\rangle,\\
&= \|u_n\|^2 - \|u_0\|^2 - 2\langle u_n - u_0\,,\, u_0\rangle.
\end{aligned}$$

Tomando esperanza y usando la Proposición 2.20 obtenemos

$$\mathtt{E}(\|u_n - u_0\|^2) = \mathcal{O}\left(\mathtt{d}_n + \mathtt{b}_n\lambda_{k_n}^{-1/2} + \mathtt{E}(\|u_n - u_0\|)w_{k_n}\right).$$

Por lo tanto

$$\mathtt{E}(\|u_n - u_0\|^2) = \mathcal{O}\left(\text{máx}\left\{\mathtt{d}_n,\, \mathtt{b}_n\lambda_{k_n}^{-1/2},\, \mathtt{E}(\|u_n - u_0\|^2)^{1/2}w_{k_n}\right\}\right).$$

Obteniendo finalmente

$$\mathtt{E}(\|u_n - u_0\|^2) = \mathcal{O}\left(\text{máx}\left\{\mathtt{d}_n,\, \mathtt{b}_n\lambda_{k_n}^{-1/2},\, w_{k_n}^2\right\}\right).$$

Lema 8.13 *Sea $u \in \mathbb{H}$ y $x_1, \ldots x_n$ una muestra aleatoria con operador de covarianza Γ con autovalores $\lambda_1 \geq \lambda_2 \geq \ldots$*

$$\|u\|_{\Gamma_n}^2 \leq \|u\|^2 \frac{4}{n}\sum_{i=1}^{n}\|x_i - m\|^2. \tag{83}$$

Por lo tanto

$$E(\|u\|_{\Gamma_n}^2) \leq 4\|u\|^2\sum_{j=1}^{\infty}\lambda_j, \tag{84}$$

$$E(\|u\|_{\Gamma_n}^4) \leq 16\|u\|^4\mathtt{E}(\|x - m\|^4). \tag{85}$$

8.14 (Lema 8.13) *Usando la propiedad de que la norma de un espacio de Hilbert satisface $\|\sum_{i=1}^n a_i\|^2 \leq n\sum\|a_i\|^2$ tenemos*

$$\begin{aligned}
\|u\|_{\Gamma_n}^2 &= \frac{1}{n}\sum_{i=1}^{n}\langle u\,,\, x_i - \overline{x}\rangle^2\,, \\
&\leq \frac{1}{n}\sum_{i=1}^{n}\|u\|^2\|x_i - \overline{x}\|^2 \\
&\leq \|u\|^2\frac{2}{n}\sum_{i=1}^{n}\left(\|x_i - m\|^2 + \|\overline{x} - m\|^2\right), \\
&\leq \|u\|^2\frac{2}{n}\sum_{i=1}^{n}\left(\|x_i - m\|^2 + \frac{1}{n}\sum_{j=1}^{n}\|x_j - m\|^2\right), \\
&= \|u\|^2\frac{4}{n}\sum_{i=1}^{n}\|x_i - m\|^2.
\end{aligned}$$

Tomando esperanza de los dos lados se obtiene 84

9. Demostraciones del Capítulo 3

9.1 (Teorema 3.2) *Como existe h_λ, tenemos que $h_\lambda + c \in \mathcal{H}$ para todo $c \in \mathbb{R}$. Por lo tanto*

$$\begin{aligned}\mathtt{E}(g(X) - h_\lambda(X))^2 + \lambda pen(h_\lambda) &= \mathtt{E}^2(g(X) - h_\lambda(X)) + \mathtt{Var}(g(X) - h_\lambda(X)) + \lambda pen(h_\lambda),\\ &\leq \mathtt{E}^2(g(X) - h_\lambda(X) - c) + \mathtt{Var}(g(X) - h_\lambda(X) - c) + \lambda pen(h_\lambda + c),\\ &= \mathtt{E}^2(g(X) - h_\lambda(X) - c) + \mathtt{Var}(g(X) - h_\lambda(X)) + \lambda pen(h_\lambda).\end{aligned}$$

Como el último término se minimiza para $c = \mathtt{E}\,(g(X) - h_\lambda(X))$ obtenemos la primera ecuación del Teorema.

Por $(C2)$ existe I cerrado tal que $1 \in I$ tal que para todo $c \in I$ vale que $ch_\lambda \in \mathcal{H}$. Además

$$\begin{aligned}E((h_\lambda(X) - g(X))^2 + \lambda pen(h_\lambda) &\leq E((ch_\lambda(X) - g(X))^2 + \lambda pen(ch_\lambda)\\ &= E(g^2(X)) - 2cE(h_\lambda(X)g(X)) + c^2\mathtt{E}(h_\lambda^2(X)) + c^q\lambda pen(h_\lambda)).\end{aligned}$$

La cuadrática se minimiza para los siguientes valores de c dependiendo de q

$$\begin{aligned}\frac{\mathtt{E}(h_\lambda(X)g(X))}{\mathtt{E}(h_\lambda^2(X))} \;&si\; q = 0,\\ \frac{\mathtt{E}(h_\lambda(X)g(X)) - \frac{\lambda}{2}pen(h_\lambda)}{\mathtt{E}(h_\lambda^2(X))} \;&si\; q = 1,\\ \frac{\mathtt{E}(h_\lambda(X)g(X))}{\mathtt{E}(h_\lambda^2(X)) + \lambda pen(h_\lambda)} \;&si\; q = 2.\end{aligned}$$

Como $ch_\lambda \in \mathcal{H}$ para todo $c \in I$ tenemos la cuadrática se minimiza necesariamente en $c = 1$. Obtenemos entonces

$$\mathtt{E}(h_\lambda^2(X)) = \begin{cases}\mathtt{E}(h_\lambda(X)g(X)) & si\; q = 0\\ \mathtt{E}(h_\lambda(X)g(X)) - \frac{\lambda}{2}pen(h_\lambda) & si\; q = 1\\ \mathtt{E}(h_\lambda(X)g(X)) - \lambda pen(h_\lambda) & si\; q = 2.\end{cases}$$

9.2 (Corolario 3.3) *Resulta*

$$\begin{aligned}\mathtt{Cov}(\varepsilon_\lambda, h_\lambda(X)) &= \mathtt{Cov}(\varepsilon, h_\lambda(X)) + \mathtt{Cov}(g(X), h_\lambda(X)) - \mathtt{Var}(h_\lambda(X)),\\ &= \mathtt{E}(g(X)h_\lambda(X)) - \mathtt{E}(g(X))\mathtt{E}(h_\lambda(X)) - E(h_\lambda^2(X)) + \mathtt{E}^2(h_\lambda(X)),\\ &= a_q\lambda pen(h_\lambda).\end{aligned}$$

Como $pen(h_\lambda) \leq pen(h_0)$ tenemos que

$$0 \leq Cov(\varepsilon_\lambda, h_\lambda(X)) \leq \lambda a_q pen(h_0).$$

9.3 (Lema 3.5) *Para esta demostración utilizamos la notación desarrollada en la sección 3.2.2. Sea L tal que $L^tL = X^tX + \rho W$ y $v \in \mathbb{R}^p$, es fácil comprobar que*

$$v^t(X^tX + \rho W)v - 2v^tX^t\tilde{y} = (L(v - \widehat{b}))^t(L(v - \widehat{b})) - \widehat{b}^t(X^tX + \rho W)\widehat{b}.$$

Como el último término no depende de v la solución de 44 es la proyección de $\widehat{b}$ sobre E bajo la métrica inducida por F.

9.4 (Proposición 3.6) *Si llamamos* $\bar{\varepsilon} = \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i$, *la predicción para una nuevo elemento aleatorio satisface*

$$\alpha^\star + \langle b^\star \,,\, x_{n+1}\rangle = \bar{Y} - \langle b^\star \,,\, \overline{x}\rangle + \langle b^\star \,,\, x_{n+1}\rangle = \alpha + \langle b \,,\, \overline{x}\rangle + \bar{\varepsilon} + \langle b^\star \,,\, x_{n+1} - \overline{x}\rangle \,.$$

Por lo tanto,

$$\alpha + \langle b \,,\, x_{n+1}\rangle - (\alpha^\star + \langle b^\star \,,\, x_{n+1}\rangle = \langle b - b^\star \,,\, x_{n+1} - \overline{x}\rangle - \bar{\varepsilon}.$$

Finalmente, tenemos

$$\begin{aligned}
\mathtt{E}\left((a + \langle b \,,\, x_{n+1}\rangle - (a^\star + \langle b^\star \,,\, x_{n+1}\rangle)\right)^2 | b^\star\Big) &= \mathtt{E}\left((\langle b - b^\star \,,\, x_{n+1} - \overline{x}\rangle - \bar{\varepsilon})^2 \,|\, b^\star\right), \\
&\leq 3\mathtt{E}\left(\langle b - b^\star \,,\, x_{n+1} - m\rangle^2 + \langle b - b^\star \,,\, m - \overline{x}\rangle^2 + \bar{\varepsilon}^2 | \, b^\star\right), \\
&\leq 3\|b - b^\star\|_\Gamma^2 + 3\|b - b^\star\|^2 \mathtt{E}(\|m - \overline{x}\|^2 | \, b^\star) + 3\mathtt{E}(\bar{\varepsilon}^2 | \, b^\star).
\end{aligned}$$

La última desigualdad se obtiene usando la Ecuación 20 y Cauchy-Schwartz para el primer y segundo término, respectivamente. Las tasas enunciadas se obtienen usando la condición (A2) y que $\|b - b^\star\|^2 = \mathcal{O}_P(1)$ *en el segundo término y que* $\bar{\varepsilon}$ *es un promedio de variables con covarianza cero.*

9.5 (Lema 9.7) *Haremos el caso* $q = 2$. *El caso general sale por inducción. Observemos que* $\|D_1 u\|^2 + \|D_2 u\|^2 = \langle u \,,\, (D_1^\star D_1 + D_2^\star D_2)u\rangle$. *Del lado izquierdo de la igualdad se desprende que* $D_1^\star D_1 + D_2^\star D_2$ *es positivo. Además resulta simétrico por ser suma de dos operadores simétricos. Como* H *es de dimensión finita esto es condición suficiente para la existencia de una raíz cuadrada* $T = (D_1^\star D_1 + D_2^\star D_2)^{1/2}$.

Lema 9.6 *Sean* $x_1, \ldots, x_n \in \mathbb{H}$ *y* $H \subseteq \mathbb{H}$ *un subespacio de dimensión* p. *Sean* $X \in \mathbb{R}^{n\times p}$ *y* $W \in \mathbb{R}^{p\times p}$ *definidos en 46. Sea* $u_0 \in \mathbb{H}$ *e* $y \in \mathbb{R}^n$ *tal que* $y_i = n^{-1/2}\langle u_0 \,,\, x_i - \overline{x}\rangle = (X|u_0|)_i$ *donde* $|u_0|$ *es la representación de* u_0 *en la base utilizada para definir* X *y* W. *Si* $w \in H$ *y* $|w| = (X^tX + \rho W)^\dagger X^t y$ *entonces*

$$w \in \operatorname*{argmin}_{u\in H} \|u_0 - u\|_{\Gamma_n}^2 + \rho \,\mathrm{Pen}(u).$$

Este lema muestra como se vinculan las soluciones de problemas de cuadrados mínimos penalizados en $\mathbb{R}^p$ con las soluciones de problemas de minimización de la norma empírica penalizada sobre un espacio de dimensión finita p.

El siguiente lema nos dice que si sumamos dos penalizaciones seguimos teniendo una penalización válida.

Lema 9.7 *Si* $D_1, D_2, \ldots, D_q$ *son operadores lineales definidos en* H *existe* $T : H \to \mathbb{H}$ *un operador lineal tal que*

$$\sum_{i=1}^{q} \|D_i u\|^2 = \|Tu\|^2.$$

Dem: 9.5

9.8 (Teorema 3.7) *Para $u \in H$ y una base del espacio podemos calcular mediante las Ecuación 47 cantidades de interés como $\|u\|^2_{\Gamma_n}$ y $Pen(u)$. Sin pérdida de generalidad podemos asumir que la base es ortonormal. Notamos $|u|$ como la representación de u en la base. A lo largo de la demostración usamos la notación del apartado 3.2.2.*

Si definimos $y_0 \in \mathbb{R}^n$ de manera que $(y_0)_i = n^{-1/2} \langle b\,,\, x_i - \overline{x} \rangle$ resulta que $(y_0)_i = \mathtt{E}_\varepsilon(\tilde{y}_i)$ y por lo tanto

$$|\mathtt{E}_\varepsilon(\widehat{b})| = \mathtt{E}_\varepsilon(|\widehat{b}|) = (X^tX + \rho W)^\dagger X^t y_0. \tag{86}$$

Por el Lema 9.6 tenemos

$$\mathtt{E}_\varepsilon(\widehat{b}) \in \operatorname*{argmin}_{u \in H} \|b - u\|^2_{\Gamma_n} + \rho\,\mathrm{Pen}(u). \tag{87}$$

Dado $u \in H$ y usando lo recién observado

$$\begin{aligned} \|b - \mathtt{E}_\varepsilon(\widehat{b})\|^2_{\Gamma_n} &\leq \|b - \mathtt{E}_\varepsilon(\widehat{b})\|^2_{\Gamma_n} + \rho\,\mathrm{Pen}(\mathtt{E}_\varepsilon(\widehat{b})), \\ &\leq \|b - u\|^2_{\Gamma_n} + \rho\,\mathrm{Pen}(u). \end{aligned}$$

Luego, obtenemos la Desigualdad 50 tomando esperanza en ambos lados y aplicando el Lema 8.13.

Para la segunda igualdad llamamos $\tilde{\varepsilon} = \tilde{y} - \mathtt{E}_\varepsilon(\tilde{y}) \in \mathbb{R}^{n\times 1}$ donde $\tilde{\varepsilon}_i = n^{-1/2}(\varepsilon_i - \bar{\varepsilon})$ y definimos $\Sigma = \mathtt{E}_\varepsilon(\tilde{\varepsilon}\tilde{\varepsilon}^t)$. Como $\Sigma_{ij} = \frac{\sigma^2}{n}(\delta_{ij} - \frac{1}{n})$ la matriz resulta semi definida positiva con exactamente un autovalor nulo y $n-1$ autovalores iguales a $\frac{n-1}{n^2}\sigma^2$. Tenemos entonces

$$\begin{aligned} \mathtt{E}_\varepsilon\left(\|\widehat{b} - \mathtt{E}_\varepsilon(\widehat{b})\|^2_{\Gamma_n}\right) &= \mathtt{E}_\varepsilon\left(\|X|\widehat{b} - \mathtt{E}_\varepsilon(\widehat{b})|\|^2\right), \\ &= \mathtt{E}_\varepsilon\left(\tilde{\varepsilon}^t X(X^tX + \rho W)^\dagger X^t X(X^tX + \rho W)^\dagger X^t \tilde{\varepsilon}\right), \\ &= Tr\left(X(X^tX + \rho W)^\dagger X^t X(X^tX + \rho W)^\dagger X^t \Sigma\right), \\ &\leq \frac{\sigma^2}{n} Tr\left(\left[(X^tX + \rho W)^\dagger X^t X\right]^2\right), \\ &\leq \frac{\sigma^2}{n} Tr\left((X^tX + \rho W)^\dagger X^t X\right). \end{aligned} \tag{88}$$

La tercera, cuarta y quinta línea se deducen de las propiedades de la traza, que se encuentran en el Apéndice, (I.1), (I.6) y (I.7) respectivamente. Obtenemos la Desigualdad 51 utilizando las Propiedades (I.6) y (I.11) en la siguiente cadena de inecuaciones

$$Tr((X^tX + \rho W)^\dagger X^t X) \leq Tr((X^tX)^\dagger X^t X) = rango(X^tX) \leq p. \tag{89}$$

Si $\rho\tau_1 > 0$ tenemos $(X^tX + \rho W)^\dagger X^tX = (X^tX + \rho W)^{-1}X^tX$. Definimos $\tilde{\chi}_{ri} = \langle \tilde{x}_i, e_r \rangle$ y para $k \in \mathbb{N}$ las matrices $X_1, X_2 \in \mathbb{R}^{n\times p}$

$$(X_1)_{ij} = \left\langle \sum_{r=1}^{k} \tilde{\chi}_{ri} e_r\,,\, \delta_j \right\rangle \quad y \quad (X_2)_{ij} = \left\langle \sum_{r=k+1}^{\infty} \tilde{\chi}_{ri} e_r\,,\, \delta_j \right\rangle.$$

Como $X = X_1 + X_2$ por la Propiedad (I.11) obtenemos

$$\mathtt{E}_\varepsilon\left(\|\widehat{b} - \mathtt{E}_\varepsilon(\widehat{b})\|^2_{\Gamma_n}\right) \leq \frac{\sigma^2}{n} \sum_{i=1}^{2} Tr((X_i^tX_i + \rho W)^{-1}X_i^tX_i).$$

Para acotar el término con $i = 1$ notemos que $X_1 = \sum_{r=1}^{k} D_r E_r$ con D_r una matriz diagonal con elementos $(D_r)_{ii} = \tilde{\chi}_{ri}$ y $(E_r)_{ij} = \langle e_r \, , \, \delta_j \rangle$ que al tener todas las filas iguales tiene rango a lo sumo uno. Concluimos que $Rango(X_1^t X_1) \leq k$ y utilizando los mismos argumentos que en la Ecuación 89 obtenemos

$$Tr((X_1^t X_1 + \rho W)^{-1} X_1^t X_1) \leq Rango(X_1^t X_1) \leq k.$$

En el término con $i = 2$ utilizamos la propiedad (I.12) con $D = W^{-1/2} X_2^t X_2^t W^{-1/2}$. Si $v_1, \ldots v_p$ son los autovectores asociados a los autovalores $\tau_1, \ldots, \tau_p$ de W. Tenemos

$$\begin{aligned} Tr((X_1^t X_1 + \rho W)^{-1} X_1^t X_1) &= \sum_{j=1}^{p} v_j^t (D + I)^{-1} D v_j, \\ &\leq l + \sum_{j=l+1}^{p} v_j^t D v_j, \\ &= l + \sum_{j=l+1}^{p} \frac{1}{\tau_j} v_j^t X_2^t X_2 v_j, \\ &\leq l + \frac{1}{\tau_{l+1}} Tr(X_2^t X_2). \end{aligned}$$

Para acotar la esperanza de la última inecuación utilizamos la notación del Teorema 1.24. Tenemos $\tilde{\chi}_{ri} = n^{-1/2}(\chi_{ri} + \langle m - \bar{x} \, , \, e_r \rangle)$ y por lo tanto

$$\begin{aligned} \sum_{i=1}^{n} \tilde{\chi}_{ri} \tilde{\chi}_{li} &= \frac{1}{n} \sum_{i=1}^{n} \chi_{ri} \chi_{li} + \langle x_i - m \, , \, e_r \rangle \langle m - \bar{x} \, , \, e_l \rangle + \langle x_i - m \, , \, e_l \rangle \langle m - \bar{x} \, , \, e_r \rangle + \langle m - \bar{x} \, , \, e_r \rangle \langle m - \bar{x} \, , \, e_l \rangle, \\ &= \frac{1}{n} \sum_{i=1}^{n} \chi_{ri} \chi_{li} - \langle m - \bar{x} \, , \, e_r \rangle \langle m - \bar{x} \, , \, e_l \rangle. \end{aligned}$$

Esta desigualdad nos permite obtener la siguiente desigualdad

$$\begin{aligned} \mathtt{E}(X_2^t X_2)_{jj} &= \mathtt{E}\left(\sum_{i=1}^{n} \left\langle \sum_{r=k+1}^{\infty} \tilde{\chi}_{ri} e_r \, , \, \delta_j \right\rangle^2 \right), \\ &= \mathtt{E}\left(\sum_{i=1}^{n} \sum_{r=k+1}^{\infty} \sum_{l=k+1}^{\infty} \tilde{\chi}_{ri} \tilde{\chi}_{li} \langle e_r \, , \, \delta_j \rangle \langle e_l \, , \, \delta_j \rangle \right), \\ &= \mathtt{E}\left(\sum_{r=k+1}^{\infty} \sum_{l=k+1}^{\infty} \langle e_r \, , \, \delta_j \rangle \langle e_l \, , \, \delta_j \rangle \left(\frac{1}{n} \sum_{i=1}^{n} \chi_{ri} \chi_{li} - \langle m - \bar{x} \, , \, e_r \rangle \langle m - \bar{x} \, , \, e_l \rangle \right) \right), \\ &= \sum_{r=k+1}^{\infty} \lambda_r \langle e_r \, , \, \delta_j \rangle^2 - \mathtt{E}\left(\sum_{r=k+1}^{\infty} \langle e_r \, , \, \delta_j \rangle \langle m - \bar{x} \, , \, e_r \rangle \right)^2, \\ &\leq \sum_{r=k+1}^{\infty} \lambda_r \langle e_r \, , \, \delta_j \rangle^2. \end{aligned}$$

Como $\{\delta_j\}$ y $\{e_r\}$ son bases ortonormales tenemos

$$\begin{aligned} Tr(\mathtt{E}(X_2^t X_2)) &= \sum_{r=k+1}^{\infty} \sum_{j=1}^{p} \lambda_r \langle e_r \, , \, \delta_j \rangle^2 \, , \\ &\leq \sum_{r=k+1}^{\infty} \lambda_r = \mathtt{s}_k . \end{aligned}$$

Finalmente, para cualquier $k, l \leq p$

$$\mathtt{E}\left(\|\widehat{b} - \mathtt{E}_\varepsilon(\widehat{b})\|_{\Gamma_n}^2 \right) \leq \frac{\sigma^2}{n} \left(k + l + \frac{1}{\rho \tau_l} \mathtt{s}_k \right).$$

9.9 (Corolario 3.8) *Continuando con la notación de la demostración del Teorema 3.7, tenemos $Y_i - \bar{Y} = \varepsilon_i - \bar{\varepsilon} + \langle b \, , \, x_i - \overline{x} \rangle$ y por lo tanto*

$$\begin{aligned} \|\widehat{b} - b\|_{\Gamma_n}^2 &= \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \bar{Y} - \left\langle \widehat{b} \, , \, x_i - \overline{x} \right\rangle - \varepsilon_i + \bar{\varepsilon} \right)^2 , \\ &= \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \bar{Y} - \left\langle \widehat{b} \, , \, x_i - \overline{x} \right\rangle \right)^2 + 2(\varepsilon_i - \bar{\varepsilon}) \left\langle \widehat{b} \, , \, x_i - \overline{x} \right\rangle + (\varepsilon_i - \bar{\varepsilon})^2 . \end{aligned}$$

Como $\widehat{b}$ es solución de 43 tenemos

$$\|\widehat{b} - b\|_{\Gamma_n}^2 + \rho \, \mathrm{Pen}(\widehat{b}) \leq \|\mathtt{E}_\varepsilon(\widehat{b}) - b\|_{\Gamma_n}^2 + \rho \, \mathrm{Pen}(\mathtt{E}_\varepsilon(\widehat{b})) + \frac{2}{n} \sum_{i=1}^{n} (\varepsilon_i - \bar{\varepsilon}) \left\langle \widehat{b} - \mathtt{E}_\varepsilon(\widehat{b}) \, , \, x_i - \overline{x} \right\rangle .$$

Utilizando la Ecuación 87 obtenemos que para cualquier $u \in H$

$$\|\widehat{b} - b\|_{\Gamma_n}^2 + \rho \, \mathrm{Pen}(\widehat{b}) \leq \|u - b\|_{\Gamma_n}^2 + \rho \, \mathrm{Pen}(u) + \frac{2}{n} \sum_{i=1}^{n} (\varepsilon_i - \bar{\varepsilon}) \left\langle \widehat{b} - \mathtt{E}_\varepsilon(\widehat{b}) \, , \, x_i - \overline{x} \right\rangle . \tag{90}$$

Si definimos, como en el demostración del Teorema 3.7, $\widetilde{\varepsilon} \in \mathbb{R}^n$ tal que $\widetilde{\varepsilon}_i = n^{-1/2}(\varepsilon_i - \bar{\varepsilon})$ y utilizamos la fórmula explícita para $\mathtt{E}_\varepsilon(\widehat{b})$ de la Ecuación 86 y las igualdades de 47 podemos obtener

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^{n} \left\langle \widehat{b} - \mathtt{E}_\varepsilon(\widehat{b}) \, , \, x_i - \overline{x} \right\rangle (\varepsilon_i - \bar{\varepsilon}) &= \sum_{i=1}^{n} n^{-1/2} (\varepsilon_i - \bar{\varepsilon}) \left\langle \widehat{b} - \mathtt{E}_\varepsilon(\widehat{b}) \, , \, n^{-1/2}(x_i - \overline{x}) \right\rangle , \\ &= \left\langle \widetilde{\varepsilon} \, , \, X(X^t X + \rho W)^{-1} X^t \widetilde{\varepsilon} \right\rangle , \\ &= \widetilde{\varepsilon}^t X (X^t X + \rho W)^{-1} X^t \widetilde{\varepsilon} . \end{aligned}$$

Como $\Sigma = \mathtt{E}(\widetilde{\varepsilon}^t \widetilde{\varepsilon})$ tiene todos sus autovalores menores a $\frac{\sigma^2}{n}$, utilizando la Propiedad (I.1) del Apéndice llegamos a

$$\mathtt{E}_\varepsilon \left(\frac{1}{n} \sum_{i=1}^{n} \left\langle \widehat{b} - \mathtt{E}_\varepsilon(\widehat{b}) \, , \, x_i - \overline{x} \right\rangle (\varepsilon_i - \bar{\varepsilon}) \right) \leq \frac{\sigma^2}{n} \, Tr\left((X^t X + \rho W)^{-1} X^t X \right).$$

Este último término aparece en 88. Por lo que volvemos a obtener las mismas cotas para su esperanza. Finalmente, tomando esperanza en ambos lados de 90 y utilizando el Lema 8.13 obtenemos los resultados deseados.

9.10 (Teorema 3.10) *En esta demostración utilizamos la notación del apartado 3.2.2 y de la demostración 9.8 para $\tilde{y}$, $\tilde{x}_i$, Σ y $\mathtt{E}_\varepsilon$. Sin pérdida de generalidad podemos asumir $b \in Nu(\Gamma)^\perp$ ya que $\langle b\,,\,\tilde{x}_i\rangle = \left\langle \Pi_{Nu(\Gamma)^\perp}(b)\,,\,\tilde{x}_i \right\rangle$.*

Utilizando una base ortonormal de H para las definiciones en 46, llamamos $S_\rho = (X^tX+\rho W)^{-1}X^t$ y $b_H = \Pi_H(b)$ nos queda

$$\|\widehat{b}\|^2 = \|S_\rho\tilde{\varepsilon}\|^2 + \|S_\rho\tilde{y}_0\|^2 + 2\,\langle S_\rho\tilde{\varepsilon}\,,\,S_\rho\tilde{y}_0\rangle\,.$$

En lo que sigue vamos a acotar la esperanza de cada uno de los términos en la suma,

$$\mathtt{E}(\langle S_\rho\tilde{\varepsilon}\,,\,S_\rho\tilde{y}_0\rangle) = \mathtt{E}(\langle S_\rho\mathtt{E}_\varepsilon(\tilde{\varepsilon})\,,\,S_\rho\tilde{y}_0\rangle) = 0.$$

Haciendo uso de las propiedades (I.1) (I.6) y (I.5) del Apéndice y que los autovalores de Σ son menores a $\frac{\sigma^2}{n}$ obtenemos

$$\begin{aligned}
\mathtt{E}_\varepsilon(\|S_\rho\tilde{\varepsilon}\|^2) &\leq \frac{\sigma^2}{n}\,Tr((X^tX+\rho W)^{-1}(X^tX+\rho W)^{-1}X^tX),\\
&\leq \frac{\sigma^2}{n}\,Tr((X^tX+\rho W)^{-2}(X^tX+\rho W)),\\
&= \frac{\sigma^2}{n}\,Tr((X^tX+\rho W)^{-1}),\\
&\leq \frac{\sigma^2}{n\rho}\,Tr(W^{-1}).
\end{aligned}$$

En el segundo sumando intercalamos con $S_\rho X|b_H| = |\operatorname{argmin}_{u\in H}\|u-b_H\|^2_{\Gamma_n}+\rho\,\mathrm{Pen}(u)|$ y obtenemos

$$\|S_\rho\tilde{y}_0\|^2 = \|S_\rho(\tilde{y}_0 - X|b_H|)\|^2 + \|S_\rho X|b_H|\|^2 + 2\,\langle S_\rho(\tilde{y}_0 - X|b_H|)\,,\,S_\rho X|b_H|\rangle\,.$$

Acotamos los términos cuadráticos y usamos Cauchy Schwartz para el producto interno. Usando las propiedades 11.1, (I.12) y (I.5) del Apéndice tenemos que,

$$\begin{aligned}
\tau_{max}\left(X(X^tX+\rho W)^{-2}X^t\right) &= \tau_{max}\left((X^tX+\rho W)^{-2}X^tX\right),\\
&\leq \tau_{max}\left((X^tX+\rho W)^{-1}\right)\tau_{max}\left((X^tX+\rho W)^{-1}X^tX\right),\\
&\leq \tau_{max}\left(X^tX+\rho W\right)^{-1},\\
&\leq (\rho\tau_1)^{-1}.
\end{aligned}$$

Por lo tanto

$$\|S_\rho\left(\tilde{y}_0 - X|b_H|\right)\|^2 \leq \frac{1}{\rho\tau_1}\|y_0 - X|b_H|\|^2 = \frac{1}{\rho\tau_1}\|b-b_H\|^2_{\Gamma_n}. \tag{91}$$

Al tomar esperanza obtenemos por el Lema 8.13

$$\mathtt{E}(\|S_\rho\left(\tilde{y}_0 - X|b_H|\right)\|^2) \leq \frac{4\mathtt{s}_0}{\rho\tau_1}\|b-b_H\|^2.$$

Usando que la base es ortonormal, la Propiedad (I.12) del Apéndice y que los autovalores de $X^tX(X^tX+\rho W)^{-2}X^tX$ son menores a uno resulta que

$$\|(X^tX+\rho W)^{-1}X^tX|b_H|\|^2 \leq \|b_H\|^2 \leq \|b\|^2. \tag{92}$$

Finalmente, usamos que $\mathtt{E}(\|x+y\|^2) \leq (\sqrt{\mathtt{E}(\|x\|^2)} + \sqrt{\mathtt{E}(\|y\|^2)})^2$ *y obtenemos*

$$\mathtt{E}(\|\widehat{b}\|^2) \leq \left(\|b\| + \sqrt{\frac{4\mathtt{s}_0}{\rho\tau_1}} \|b - b_H\| \right)^2 + \frac{\sigma^2}{n\rho} Tr\left(W^{-1}\right).$$

Para acotar

$$\mathtt{E}\left(\|\widehat{b}\|^4\right) \leq 8\|S_\rho \widetilde{\varepsilon}\|^4 + 8\|S_\rho \widetilde{y}_0\|^4,$$

tenemos que $\mathtt{E}(\|S_\rho\widetilde{\varepsilon}\|^4) = \mathtt{Var}\left(\|S_\rho\widetilde{\varepsilon}\|^2\right) + \mathtt{E}\left(\|S_\rho\widetilde{\varepsilon}\|^2\right)^2$. *Por la ley de la Varianza Total*

$$\mathtt{Var}(\|S_\rho\widetilde{\varepsilon}\|^2) = \mathtt{E}\left(\mathtt{Var}(\|S_\rho\widetilde{\varepsilon}\|^2|S_\rho)\right) + \mathtt{Var}\left(\mathtt{E}(\|S_\rho\widetilde{\varepsilon}\|^2)|S_\rho\right).$$

Si llamamos $Q_\rho = S_\rho^t S_\rho$ *tenemos*

$$\mathtt{E}(\mathtt{Var}(\|S_\rho\widetilde{\varepsilon}\|^2)|S_\rho) = \mathtt{E}(\mathtt{Var}(\widetilde{\varepsilon}^t Q_\rho \widetilde{\varepsilon})|Q_\rho) = \mathtt{E}(\,Tr\,(Q_\rho \Sigma Q_\rho \Sigma)) \leq \frac{\sigma^4}{n^2} E(\,Tr\left(Q_\rho^2\right)).$$

Con argumentos similares a los que usamos para obtener la cota de $\mathtt{E}(\|S_\rho\widetilde{\varepsilon}\|^2)$ *tenemos*

$$\begin{aligned} Tr\left(Q_\rho^2\right) &= Tr\left((X^tX + \rho W)^{-2} X^tX (X^tX + \rho W)^{-2} X^tX\right), \\ &\leq Tr\left((X^tX + \rho W)^{-2}\right), \\ &\leq \frac{1}{\rho^2} Tr\left(W^{-2}\right), \\ &\leq \left(\frac{Tr(W^{-1})}{\rho}\right)^2, \end{aligned}$$

donde la última desigualdad vale porque la traza es la suma de los autovalores que son todos positivos.

En Popoviciu (1933) se muestra que si Y *es una variable aleatoria tal que* $m \leq Y \leq M$ *con probabilidad* 1 *entonces*

$$\mathtt{Var}(Y) \leq \frac{(M-m)^2}{4}.$$

Como $0 \leq \mathtt{E}(\|S_\rho\widetilde{\varepsilon}\|^2|S_\rho) \leq \frac{\sigma^2}{n\rho} Tr(W^{-1})$ *obtenemos*

$$\mathtt{Var}(\mathtt{E}(\|S_\rho\widetilde{\varepsilon}\|^2|S_\rho)) \leq \frac{1}{4}\left(\frac{\sigma^2}{n\rho} Tr(W^{-1})\right)^2.$$

Concluimos

$$\mathtt{E}(\|S_\rho\widetilde{\varepsilon}\|^4) = \mathtt{Var}(\|S_\rho\widetilde{\varepsilon}\|^2) + \mathtt{E}^2(\|S_\rho\widetilde{\varepsilon}\|^2) \leq \frac{9}{4}\left(\frac{\sigma^2}{n\rho} Tr(W^{-1})\right)^2.$$

De la Desigualdad 92

$$\mathtt{E}(\|S_\rho\tilde{y}_0\|^4) \leq 8\mathtt{E}(\|S_\rho\tilde{y}_0 - X|b_H|\|^4) + 8\mathtt{E}(\|S_\rho X|b_H|\|^4) \leq 8\mathtt{E}(\|S_\rho(\tilde{y}_0 - X|b_H|)\|^4) + 8\|b\|^4.$$

Finalmente, por la desigualdad 91 y el Lema 8.13

$$\mathtt{E}\left(\|S_\rho(\tilde{y}_0 - X|b_H)|\|^4\right) \leq \frac{1}{\rho^2\tau_1^2}\mathtt{E}\left(\|b - b_H\|^4_{\Gamma_n}\right) \leq 16\mathtt{s}_0^4\left(\frac{\|b-b_H\|^2}{\rho\tau_1}\right)^2 \mathtt{E}\left(\|x-m\|^4\right).$$

9.11 (Teorema 3.13) *Para aligerar la notación en esta demostración omitiremos los módulos para hablar de la representación en una base de $\widehat{b}$, $\widetilde{b}$ y u. Mismo notaremos como E al conjunto $T(E)$ definido en 49.*

Para obtener 57 hacemos uso de la siguiente desigualdad que vale para cualquier $u \in E$,

$$\|\widetilde{b} - u\|^2_{\Gamma_n} \leq \|\widehat{b} - u\|^2_{\Gamma_n} + 2\rho Pen(\widehat{b} - u). \tag{93}$$

Este resultado lo obtenemos utilizando las ideas del trabajo de Meyer (2012) que está planteado para un tipo de conjuntos convexos particulares y para una penalización específica enmarcado en un problema de regresión no paramétrica restringida.

Sean X y W definidos en 46 con $W = D^tD$ inversible por hipótesis. Definimos $\hat{\mu} = X\widehat{b}$, $\widetilde{\mu} = X\widetilde{b}$ y $\mu = Xu$ para $u \in E$ y tenemos por las igualdades en 47 que la Desigualdad 93 es equivalente a

$$\|\widetilde{\mu} - \mu\|^2 \leq \|\widehat{\mu} - \mu\|^2 + 2\rho\|D(\widehat{b} - u)\|^2.$$

Sea $L \in \mathbb{R}^{p\times p}$ tal que $L = (X^tX + \rho W)^{1/2}$. Por hipótesis L es inversible. Consideremos el problema de minimización equivalente a 49 al que se llega mediante la transformación $\phi = Lv$, $z = L^{-1}X^ty$ y $L(E)$ la imagen de L restringido al conjunto E.

$$\begin{aligned} \underset{\phi\in\mathbb{R}^p}{\operatorname{argmin}} \quad & \|\phi - z\|^2 \\ r.a. \quad & \phi \in L(E). \end{aligned} \tag{94}$$

Para justificar la equivalencia notar que

$$v^t\left(X^tX + \rho W\right)v - 2y^tXv = \phi^t\phi - 2z^t\phi = (\phi - z)^t(\phi - z) - z^tz.$$

Como z^tz es constante las funciones objetivos son equivalentes. Esto también sucede con la restricción. Es decir $v \in E$ si y solo si $L^{-1}\phi \in E$ si y solo si $\phi \in L(E)$.

Como $L(E)$ es convexo la solución del problema es única. Si $\widetilde{\phi}$ es solución de 94 y $\widetilde{b}$ de 49 ambas cumplen

$$\widetilde{b} = L^{-1}\widetilde{\phi}.$$

Por la Proposición 1.1 de Bertsekas (2009), $\widetilde{\phi}$ es solución de 94 si y solo si

$$(z - \widetilde{\phi})^t(\phi - \widetilde{\phi}) \leq 0 \quad \forall \phi \in L(E).$$

Esta desigualdad en término del problema original implica que

$$(y - \widetilde{\mu})^t(\mu - \widetilde{\mu}) \leq \rho\widetilde{b}W(u - \widetilde{b}) \quad \forall\, \mu = Xu, \quad u \in E, \tag{95}$$

pues

$$\begin{aligned} 0 \geq (z - \widetilde{\phi})^t(\phi - \widetilde{\phi}) &= (L^{-1}X^ty - L^t\widetilde{b})^tL^t(u - \widetilde{b}), \\ &= y^tX(u - \widetilde{b}) - \widetilde{b}^tLL^t(u - \widetilde{b}), \\ &= y^tX(u - \widetilde{b}) - \widetilde{b}^tX^tX(u - \widetilde{b}) - \rho\widetilde{b}^tW(u - \widetilde{b}), \\ &= (y - X\widetilde{b})^tX(u - \widetilde{b}) - \rho\widetilde{b}^tW(u - \widetilde{b}), \\ &= (y - \widetilde{\mu})^t(\mu - \widetilde{\mu}) - \rho\widetilde{b}^tW(u - \widetilde{b}). \end{aligned}$$

Además por condiciones de optimalidad de la solución no restringida

$$(y-\widehat{\mu})^t\mu = \rho\widehat{b}^t W u \quad \forall \mu = Xu \qquad . \tag{96}$$

Utilizando estas propiedades obtenemos la siguiente desigualdad para cualquier $\mu = Xu$ con $u \in E$.

$$\begin{aligned}
\|\widehat{\mu}-\mu\|^2 &= \|\widehat{\mu}-\widetilde{\mu}\|^2 + \|\widetilde{\mu}-\mu\|^2 + 2(\widehat{\mu}-\widetilde{\mu})^t(\widetilde{\mu}-\mu),\\
&= \|\widehat{\mu}-\widetilde{\mu}\|^2 + \|\widetilde{\mu}-\mu\|^2 - 2(y-\widehat{\mu})^t(\widetilde{\mu}-\mu) + 2(y-\widetilde{\mu})^t(\widetilde{\mu}-\mu),\\
&\geq \|\widehat{\mu}-\widetilde{\mu}\|^2 + \|\widetilde{\mu}-\mu\|^2 + 2\rho(\widetilde{b}-\widehat{b})^t W(\widetilde{b}-u).
\end{aligned}$$

Para la última desigualdad notamos que $\widetilde{\mu}-\mu = X(\widetilde{b}-u)$ junto con la Ecuación 96 para obtener $-2(y-\widehat{\mu})^t(\widetilde{\mu}-\mu) = -2\rho\,\widehat{b}W(\widetilde{b}-u)$. Como $u \in E$ usamos la desigualdad 95 para obtener $2(y-\widetilde{\mu})^t(\widetilde{\mu}-\mu) \geq 2\rho\widetilde{b}W(\widetilde{b}-u)$.

Llegamos a

$$\|\widehat{\mu}-\mu\|^2 - \|\widetilde{\mu}-\mu\|^2 \geq \|\widetilde{\mu}-\widehat{\mu}\|^2 + 2\rho\|D(\widetilde{b}-\widehat{b})\|\left(\|D(\widetilde{b}-\widehat{b})\| - \|D(\widehat{b}-u)\|\right).$$

Existen dos posibilidades, o bien $\|D(\widehat{b}-u)\| < \|D(\widetilde{b}-\widehat{b})\|$ y por lo tanto $\|\widetilde{\mu}-\mu\|^2 \leq \|\widehat{\mu}-\mu\|^2$ o bien $\|D(\widehat{b}-u)\| \geq \|D(\widetilde{b}-\widehat{b})\|$ donde queda

$$\|\widehat{\mu}-\mu\|^2 - \|\widetilde{\mu}-\mu\|^2 \geq \|\widehat{\mu}-\widetilde{\mu}\|^2 - 2\rho\|D(\widetilde{b}-\widehat{b})\|\|D(\widehat{b}-u)\| \geq \|\widehat{\mu}-\widetilde{\mu}\|^2 - 2\rho\|D(\widehat{b}-u)\|^2.$$

Cualquiera de las dos implica

$$\|\widetilde{\mu}-\mu\|^2 \leq \|\widehat{\mu}-\mu\|^2 + 2\rho\|D(\widehat{b}-u)\|^2.$$

Obtenemos la Desigualdad 57 utilizando 93 puesto que para cualquier $u \in E$

$$\begin{aligned}
\|\widetilde{b}-b\|^2_{\Gamma_n} &\leq 2\|\widetilde{b}-u\|^2_{\Gamma_n} + 2\|u-b\|^2_{\Gamma_n},\\
&\leq 2\|\widehat{b}-u\|^2_{\Gamma_n} + 2\|b-u\|^2_{\Gamma_n} + 4\rho\,\mathrm{Pen}(\widehat{b}-u),\\
&\leq 4\|\widehat{b}-b\|^2_{\Gamma_n} + 6\|b-u\|^2_{\Gamma_n} + 4\rho\,\mathrm{Pen}(\widehat{b}-u).
\end{aligned}$$

Para la Desigualdad 58 usamos que $\|\widetilde{b}\|^2 \leq 2\|\widehat{b}\|^2 + 2\|\widehat{b}-\widetilde{b}\|^2$ y acotamos el segundo término utilizando el Lema 3.5 en la segunda desigualdad.

$$\begin{aligned}
\rho\tau_1\|\widehat{b}-\widetilde{b}\|^2 &= \rho\tau_1\|L^{-1}L(\widehat{b}-\widetilde{b})\|^2\\
&\leq \|L(\widehat{b}-\widetilde{b})\|^2\\
&= \|\widehat{b}-\widetilde{b}\|^2_{\Gamma_n} + \rho\,\mathrm{Pen}(\widehat{b}-\widetilde{b})\\
&\leq \|\widehat{b}-u\|^2_{\Gamma_n} + \rho\,\mathrm{Pen}(\widehat{b}-u).
\end{aligned}$$

9.12 (Proposición 3.16) *Por el Teorema 3.7 y que D es un operador acotado tenemos tenemos*

$$\mathtt{E}\left(\|\widehat{b}_n-b\|^2_{\Gamma_n}\right) = \mathcal{O}\left(\|b-b_{H_n}\|^2 + \rho_n\|b\|^2 + \frac{p_n}{n}\right).$$

Para las desigualdades del Teorema 3.10 obtenemos de la coercividad de D que $\sum_{j=1}^{p_n}\tau_j^{-1} = \mathcal{O}(p_n)$.

9.13 (Proposición 3.20) *Como D es inyectivo y $\operatorname{Pen}(u) \leq \rho\|D\|^2\|b\|^2$ podemos utilizar la Ecuación 59. Si K es una cota superior para $\mathcal{C}$ resulta $\|\widetilde{b}\|^4 \leq K^4$.*

9.14 (Proposición 3.26) *Por continuidad de las derivadas, existe $c > 0$ tal que si $t \in [a,b]$ entonces $L_{\vec{a}}(f)(t) > 2c$. Por el Teorema 3.22 tenemos para una partición*

$$\|L(f-\pi f)\|_\infty \leq \sum_{r=1}^{k-1} |a_r| \|(f-\pi f)^{(r)}\|_\infty \leq \|f^{(k)}\|_\infty D_k \sum_{r=0}^{k-1} |a_r| h^{k-r}. \tag{97}$$

Como

$$L(\pi f)(t) = L(f)(t) + L(\pi f)(t) - L(f)(t) \geq 2c - \|L(\pi f)(t) - L(f)(t)\|_\infty,$$

tomando $h_0 = c\left(k\|f^{(k)}\|_\infty D_k \operatorname{máx}\{|a_r|\}\right)^{-1}$ resulta $L(\pi f)(t) > 0$.

El caso con $\|f^{(k)}\|_\infty = 0$ resulta que $\pi(f) = f$ y el resultado vale trivialmente.

9.15 (Proposición 3.27) *Sea Π una partición con calibre h. Construiremos una función $g \in \mathcal{L}_{\vec{a}}$ que satisface lo buscado para una constante que no dependa de Π. Por la Ecuación 97 de la demostración anterior tenemos que si Π es una partición de calibre h*

$$\|L(f-\pi f)\|_\infty \leq \|f^{(k)}\|_\infty \sum_{r=0}^{k-1} |a_r| D_k h^{k-r}.$$

Sea $D = k\|f^{(k)}\|_\infty D_k \operatorname{máx}_r\{|a_r|\}$, esta constante es independiente de Π y vale que

$$\|L(f) - L(\pi f)\|_\infty \leq D h^{k-r_1}.$$

Llamemos $r_0 = \operatorname{mín}_r\{|a_k|\}$. Construimos la siguiente función con el objetivo de encontrar un elemento de $\mathcal{L}_{\vec{a}} \cap S^k_\Pi$ cerca de πf

$$g(t) = \pi f(t) + D h^{k-r_1} \frac{(t-a)^{r_0}}{a_{r_0} r_0!},$$

Por construcción $g \in S^k_\Pi$, veamos que además $g \in \mathcal{L}_{\vec{a}}$.

Como $L_{\vec{a}}((t-a)^{r_0}) = r_0! a_{r_0}$ obtenemos que

$$\begin{aligned} L_{\vec{a}}(g)(t) &= L(\pi f)(t) + D h^{k-r_1}, \\ &\geq L(\pi f)(t) + D h^{k-r_1} - L(f)(t), \\ &\geq D h^{k-r_1} - \|L(f) - L(\pi f)\|_\infty \geq 0. \end{aligned}$$

Por último tomando $D' = \frac{D}{a_{r_0} r_0!}$ tenemos que $\|f-g\| \leq \|f - \pi f\| + D'(b-a)^{r_0} h^{k-r_0}$.

9.16 (Proposición 3.28) *Consideramos primero el caso $b \in \mathcal{L}_{\vec{a},0}$. Sabemos de la Proposición 3.26 que existe p_0 tal que si $p \geq p_0$ $\pi b \in \mathcal{L}_{\vec{a}}$. Utilizando las Desigualdades 59 y 60, considerando que $\operatorname{Pen}_n(\pi b) \leq 2(\operatorname{Pen}_n(b) + \operatorname{Pen}_n(b - \pi b)) = \mathcal{O}(1)$ y de las tasas del estimador sin restricción dadas en 62 obtenemos el resultado.*

El caso con $b \in \mathcal{L}_{\vec{a}}$ sigue de la Proposición 3.27 y la Proposición 3.19.

10. Demostraciones del Capítulo 4

10.1 (Teorema 4.2) *En la demostración omitiremos el subíndice de $b_n^\star$ para descargar la notación.*

Notemos que para cualquier $\delta > 0$

$$\left\{|\hat{\theta} - \theta_0| > \delta\right\} \subseteq \left\{PL(\theta_0, b^\star, P_n) \geq \inf_{|\theta-\theta_0|>\delta} PL(\theta, b^\star, P_n)\right\}.$$

Si $\theta > \theta_0 + \delta$ por la Ecuación 67 tenemos que $\ell(\theta_0, b, P) = \ell(\theta, b, P)$ y por lo tanto

$$\begin{aligned} PL(\theta_0, b^\star, P_n) - PL(\theta, b^\star, P_n) &= \ell(\theta_0, b^\star, P_n) - \ell(\theta, b^\star, P_n) + \omega_n(pen(\theta_0) - pen(\theta)), \\ &= \ell(\theta_0, b^\star, P_n) - \ell(\theta_0, b, P) + \ell(\theta, b, P) - \ell(\theta, b^\star, P_n) + \omega_n(pen(\theta_0) - pen(\theta)), \\ &\leq 2S_n + \omega_n(pen(\theta_0) - pen(\theta_0 + \delta)). \end{aligned}$$

Es decir

$$si \quad PL(\theta_0, b^\star, P_n) - PL((\theta, b^\star, P_n) \geq 0 \implies 2S_n + \omega_n(pen(\theta_0) - pen(\theta_0 + \delta) \geq 0.$$

Por lo tanto

$$P\left(PL(\theta_0, b^\star, P_n) \geq \inf_{\theta>\theta_0+\delta} PL(\theta, b^\star, P_n)\right) \leq P\left(2S_n \geq \omega_n(pen(\theta_0 + \delta) - pen(\theta_0))\right). \tag{98}$$

Como $\mathtt{a}_n = o(\omega_n)$ resulta que $2{\omega_n}^{-1}S_n = o(1)$. Además $pen(\theta_0 + \delta) - pen(\theta_0) > 0$ y por lo tanto existe $n_1 = n_1(\delta, \varepsilon)$ tal que si $n \geq n_1$

$$P\left(PL(\theta_0, b^\star, P_n) \geq \inf_{\theta>\theta_0+\delta} PL(\theta, b^\star, P_n)\right) < \frac{\varepsilon}{2}.$$

Cuando $\theta < \theta_0 - \delta$ tenemos

$$\begin{aligned} PL(\theta_0, b^\star, P_n) - PL(\theta, b^\star, P_n) &= \ell(\theta_0, b^\star, P_n) - \ell(\theta, b^\star, P_n) + \omega_n(pen(\theta_0) - pen(\theta)), \\ &= \ell(\theta_0, b^\star, P_n) - \ell(\theta_0, b, P) + \ell(\theta, b, P) - \ell(\theta, b^\star, P_n), \\ &\quad + \ell(\theta_0, b, P) - \ell(\theta, b, P) + \omega_n(pen(\theta_0) - pen(\theta)), \\ &\leq 2S_n - \alpha(\delta) + \omega_n(pen(\theta_0) - pen(\theta_0 - \delta)), \\ &\leq 2S_n - \alpha(\delta) + \omega_n pen(\theta_0). \end{aligned}$$

Obtenemos entonces

$$P\left(PL(\theta_0, b^\star, P_n) \geq \inf_{\theta<\theta_0-\delta} PL(\theta, b^\star, P_n)\right) \leq P\left(2S_n + \omega_n pen(\theta_0) \geq \alpha(\delta)\right). \tag{99}$$

Siendo $\mathtt{a}_n = o(\omega_n)$ y $\omega_n = o(1)$ tenemos que $2S_n + \omega_n pen(\theta_0) = o(1)$. Al mismo tiempo la condición (B1) implica que $\alpha(\delta) > 0$. Por lo tanto, existe $n_2 = n_2(\delta, \varepsilon)$ tal que

$$P\left(PL(\theta_0, b^\star, P_n) \geq \inf_{\theta<\theta_0-\delta} PL(\theta, b^\star, P_n)\right) < \frac{\varepsilon}{2}.$$

Finalmente para $\delta, \varepsilon > 0$ tomando $n_0 = \max\{n_1, n_2\}$ tenemos que para todo $n \geq n_0$

$$P\left(|\hat{\theta}_n - \theta_0| > \delta\right) \leq \varepsilon.$$

Para probar la Ecuación 74 con $\delta_n = \mathtt{a}_n^{1/(q+1)}$ y $\omega_n = \mathtt{a}_n^{q/(q+1)}$ tenemos que ver que dado $\varepsilon > 0$ existe $K > 0$ y $n_0 \in \mathbb{N}$ tales que para todo $n \geq n_0$

$$P\left(|\hat{\theta}_n - \hat{\theta}_0| > K\delta_n\right) \leq \varepsilon.$$

Bajo las Hipótesis (P1) y (P2) obtenemos continuando la Desigualdad 98

$$\begin{aligned} P\left(PL(\theta_0, b^\star, P_n) \geq \inf_{\theta > \theta_0 + K\delta_n} PL(\theta, b^\star, P_n)\right) &\leq P\left(2S_n \geq C_b \omega_n K \delta_n\right), \\ &= P\left(S_n \geq \frac{C_b K}{2} \mathtt{a}_n\right). \end{aligned}$$

Luego, dado $\varepsilon > 0$, como $S_n = \mathcal{O}_P(\mathtt{a}_n)$ existe $n_1(\varepsilon)$ tal que si K es lo suficientemente grande, para todo $n \geq n_1$

$$P\left(S_n \geq \frac{C_b K}{2} \mathtt{a}_n\right) \leq \varepsilon/2.$$

De la Hipótesis (B2), la Ecuación 99 queda

$$\begin{aligned} P\left(PL(\theta_0, b^\star, P_n) \geq \inf_{\theta < \theta_0 - K\delta_n} PL(\theta, b^\star, P_n)\right) &\leq P\left(2S_n + \omega_n pen(\theta_0) \geq C_b K^q \delta_n^q\right), \\ &= P\left(S_n \geq \frac{1}{2}(C_b K^q - pen(\theta_0))\omega_n\right). \end{aligned}$$

Como $S_n = o_P(\omega_n)$ si K es suficientemente grande como para que $C_b K^q - pen(\theta_0) > 0$, existe n_2 tal que para todo $n \geq n_2$ resulta

$$P\left(S_n \geq \frac{1}{2}(C_b K^q - pen(\theta_0))\omega_n\right) \leq \varepsilon/2.$$

Finalmente para $n \geq \max\{n_1, n_2\}$ y K suficientemente grande resulta

$$P(|\hat{\theta}_n - \hat{\theta}_0| \geq K \mathtt{a}_n^{1/(q+1)}) \leq \varepsilon.$$

10.2 (Proposición 4.3) *Notemos que*

$$\begin{aligned} \ell(\theta, b^\star, P_n) &= P_n\left(Y - P_n Y - \langle b_\theta^\star\,,\, x - P_n x\rangle\right)^2 \\ &= P_n\left(\varepsilon - P_n\varepsilon + \langle b - b_\theta^\star\,,\, x - P_n x\rangle\right)^2 \\ &= P_n\left(\varepsilon - P_n\varepsilon + \langle b_\theta - b_\theta^\star\,,\, x - P_n x\rangle + \langle b - b_\theta\,,\, x - P_n x\rangle\right)^2. \end{aligned}$$

Por la Ecuacíon 67 resulta

$$\sup_{\theta\in[0,T]} |\ell(\theta, b^\star, P_n) - \ell(\theta_0, b, P)| \leq \quad \left|P_n(\varepsilon - P_n\varepsilon)^2 - P(\varepsilon^2)\right| \tag{100}$$

$$+ \sup_{\theta\in[0,T]} \left|P_n \langle b - b_\theta\,,\, x - P_n x\rangle^2 - P(\langle b - b_\theta\,,\, x - Px\rangle^2\right| \tag{101}$$

$$+ \sup_{\theta\in[0,T]} P_n \langle b_\theta - b_\theta^\star\,,\, x - P_n x\rangle^2 \tag{102}$$

$$+ 2 \sup_{\theta\in[0,T]} |P_n(\varepsilon - P_n\varepsilon) \langle b - b_\theta\,,\, x - P_n x\rangle| \tag{103}$$

$$+ 2 \sup_{\theta\in[0,T]} |P_n(\varepsilon - P_n\varepsilon) \langle b_\theta - b_\theta^\star\,,\, x - P_n x\rangle| \tag{104}$$

$$+ 2 \sup_{\theta\in[0,T]} |P_n(\langle b - b_\theta\,,\, x - P_n x\rangle \langle b_\theta - b_\theta^\star\,,\, x - P_n x\rangle|\,. \tag{105}$$

Tenemos gracias al Teorema Central del Límite

$$|P_n(\varepsilon - P_n\varepsilon)^2 - P(\varepsilon^2)| = \mathcal{O}_P(n^{-1/2}).$$

Como $\|b - b_\theta\|^2 \leq \|b\|^2$ *la expresión en 101 se acota por*

$$\sup_{\theta\in[0,T]} |\|b - b_\theta\|_{\Gamma_n}^2 - \|b - b_\theta\|_\Gamma^2| \leq \|b\|^2 \sup_{\|u\|\leq 1} |\|u\|_{\Gamma_n}^2 - \|u\|_\Gamma^2| = \mathcal{O}_P(n^{-1/2}).$$

Por el Lema 8.13 tenemos que $\sup_{\theta\in[0,T]} \|b - b_\theta\|_{\Gamma_n}^2 = \mathcal{O}_P(\|b\|^2)$*. Por lo tanto, los términos en 104 y 105 quedan acotados respectivamente por*

$$\left(P_n(\varepsilon - P_n\varepsilon)^2\right)^{1/2} \sup_{\theta\in[0,T]} \|b_\theta - b_\theta^\star\|_{\Gamma_n} = \mathcal{O}_P(\mathtt{a}_n),$$

$$\sup_{\theta\in[0,T]} \|b - b_\theta\|_{\Gamma_n} \sup_{\theta\in[0,T]} \|b_\theta - b_\theta^\star\|_{\Gamma_n} = \mathcal{O}_P(\mathtt{a}_n).$$

Si bien la Ley de los grandes números nos da convergencia para cada θ *fijo en 103 necesitamos un argumento de uniformidad sobre el intervalo* $[0, \theta_0]$*. Utilizaremos un resultado estándar de la Teoría de Procesos Empíricos.*

Para $j = 0, 1$ *consideremos las clases de funciones* $\mathcal{F}_j = \{g_{j,\theta} : \theta \in [0, \theta_0)\}$ *con* $g_{j,\theta}(\varepsilon, x) = \varepsilon^j \langle b - b_\theta\,,\, x - Px\rangle$ *y notemos que* $Pg_{j,\theta}(\varepsilon, x) = 0$ *para todo* $\theta \in [0, \theta_0)$*. Cada una de estas clases es Lipchitz en el sentido de que*

$$|g_{j,\theta_1}(\varepsilon, x) - g_{j,\theta_2}(\varepsilon, x)| \leq |\varepsilon^j| \|b\|_\infty \|x - Px\|^2 \sqrt{|\theta_1 - \theta_2|}.$$

Además, tenemos que $F_j(\varepsilon, x) = \theta_0 |\varepsilon^j| \|b\|_\infty \|x - Px\|^2$ *es una envolvente con esperanza cuadrada finita siempre que* $P\|x\|^4 < +\infty$*. Por el Teorema 2.7.11 de Van Der Vaart (1996) el número de bracketing de estas clases esta acotado por el número de cobertura del intervalo* $[0, \theta_0)$*, es decir*

$$N_{[]}(2\varepsilon\|F_j\|, \mathcal{F}_j, L_2(P)) \leq C_1(\|F_j\|) N(\varepsilon, (0, \theta_0), d),$$

con $d(\theta_1,\theta_2)=\sqrt{|\theta_1-\theta_2|}$ *y* $C(\|F_j\|)$ *constantes finitas que dependen de la norma de la envolvente. Como* $N(\varepsilon,(0,\theta_0),d)=N(\varepsilon^2,[0,\theta_0),\|\cdot\|)=\frac{\theta_0}{\varepsilon^2}$ *tenemos que*

$$\int_0^1\sqrt{\log N_{[]}(2\varepsilon\|F_j\|,\mathcal{F}_|,L_2(P))}\,d\varepsilon<C_2(\|F_j\|)\int_0^1\sqrt{\log N(\varepsilon,(0,\theta_0),d)}\,d\varepsilon<+\infty.$$

Esto implica que tanto $\mathcal{F}_0$ *y* $\mathcal{F}_1$ *son clases Donsker y tenemos que*

$$\sup_{\theta\in[0,\theta_0)}\left|P_n\varepsilon^j\left\langle b-b_\theta\,,\,x-P_nx\right\rangle\right|=\sup_{g_\theta\in\mathcal{F}_j}|P_ng_{\theta,j}(\varepsilon,x)|=\mathcal{O}_P\left(n^{-1/2}\right).$$

Como $b_\theta=b$ *si* $\theta\geq\theta_0$ *conlcuímos*

$$\sup_{\theta\in[0,T]}\left|P_n\varepsilon^j\left\langle b-b_\theta\,,\,x-P_nx\right\rangle\right|=\mathcal{O}_P\left(n^{-1/2}\right).$$

10.3 (Teorema 4.6) *Con el fin de aligerar la notación omitimos el indice que indica el tamaño de muestran para* $\widehat{b}_n[\theta]$*. Como* $Y_i-\bar{Y}=\varepsilon_i-\bar{\varepsilon}+\langle b\,,\,x_i-\overline{x}\rangle$ *tenemos para* $\theta\in[0,T]$

$$\begin{aligned}\|\widehat{b}[\theta]-b\|_{\Gamma_n}^2&=\frac{1}{n}\sum_{i=1}^n\left(Y_i-\bar{Y}-\left\langle\widehat{b}[\theta]\,,\,x_i-\overline{x}\right\rangle-\varepsilon_i+\bar{\varepsilon}\right)^2,\\&=\frac{1}{n}\sum_{i=1}^n\left(Y_i-\bar{Y}-\left\langle\widehat{b}[\theta]\,,\,x_i-\overline{x}\right\rangle\right)^2+2(\varepsilon_i-\bar{\varepsilon})\left\langle\widehat{b}[\theta]\,,\,x_i-\overline{x}\right\rangle+(\varepsilon_i-\bar{\varepsilon})^2.\end{aligned}$$

Como $\widehat{b}[\theta]$ *se define por la Ecuación 71 resulta*

$$\|\widehat{b}[\theta]-b\|_{\Gamma_n}^2+\rho\|\widehat{b}[\theta]\|^2\leq\|\mathtt{E}_\varepsilon(\widehat{b}[\theta])-b\|_{\Gamma_n}^2+\rho\|\mathtt{E}_\varepsilon(\widehat{b}[\theta])\|^2+\frac{2}{n}\sum_{i=1}^n(\varepsilon_i-\bar{\varepsilon})\left\langle\widehat{b}[\theta]-\mathtt{E}_\varepsilon(\widehat{b}[\theta]))\,,\,x_i-\overline{x}\right\rangle.$$

De la Ecuación 87 en la demostración 9.8 tenemos que la esperanza condicional minimiza la norma empírica penalizada sobre todos los elementos de H_θ*. En particular para la proyección ortogonal de* b *sobre* H_θ *que notamos* $\Pi_{H_\theta}(b)$*.*

$$\|\widehat{b}[\theta]-b\|_{\Gamma_n}^2+\rho\|\widehat{b}[\theta]\|^2\leq\|\Pi_{H_\theta}(b)-b\|_{\Gamma_n}^2+\rho\|\Pi_{H_\theta}(b)\|^2+\frac{2}{n}\sum_{i=1}^n(\varepsilon_i-\bar{\varepsilon})\left\langle\widehat{b}[\theta]-\mathtt{E}_\varepsilon(\widehat{b}[\theta])\,,\,x_i-\overline{x}\right\rangle.\quad(106)$$

Por el Lema 8.13 y que $\Pi_{H_\theta}(b)=\Pi_{H_\theta}(b_\theta)$ *es idénticamente nula a partir de* θ *tenemos*

$$\|\Pi_{H_\theta}(b)-b\|_{\Gamma_n}^2+\rho\|\Pi_{H_\theta}(b)\|^2\leq\|\Pi_{H_\theta}(b)-b\|^2 4n^{-1}\sum_{i=1}^n\|x_i-\overline{x}\|^2+\rho\|b\|^2,\quad(107)$$

$$=\frac{4}{n}\sum_{i=1}^n\|x_i-\overline{x}\|^2\left(\|\Pi_{H_\theta}(b_\theta)-b_\theta\|^2+\delta\|b\|^2\right)+\rho\|b\|^2.\quad(108)$$

Nos enfocamos ahora en acotar el último término de la Desigualdad 106. Sea $\{g_1,\ldots,g_p\}$ *una base ortonormal de* H *tenemos definida* $\{f_1,\ldots,f_p\}$ *una base ortonormal de* H_θ *mediante la Ecuación*

70. Definimos entonces $X_\theta \in \mathbb{R}^{n\times p}$

$$\begin{aligned}(X_\theta)_{ij} &= n^{-1/2}\int_0^\theta f_j(s)(x_i(s)-\overline{x}(s))\,ds,\\ &= n^{-1/2}\theta^{1/2}\int_0^\theta g_j(\frac{s}{\theta})(x_i(s)-\overline{x}(s))\,ds,\\ &= n^{-1/2}\theta^{3/2}\int_0^1 g_j(t)(x_i(\theta t)-\overline{x}(\theta t))\,dt.\end{aligned}$$

Luego de la Sección 3.2.2 con $W = Id$ *resulta que los coeficientes en la base* $\{f_1,\dots f_p\}$ *para* $\widehat{b}[\theta]$ *vienen dados por*

$$|\widehat{b}[\theta]| = (X_\theta^t X_\theta + \rho I)^{-1}X_\theta^t\tilde{y}.$$

Si definimos, como en el demostración del Teorema 3.7, $\widetilde{\varepsilon}\in\mathbb{R}^n$ *tal que* $\widetilde{\varepsilon}_i = n^{-1/2}(\varepsilon_i-\bar{\varepsilon})$ *y utilizamos la fórmula explícita para* $\mathtt{E}_\varepsilon(\widehat{b}[\theta])$ *de la Ecuación 86 y las igualdades de 47 podemos obtener, usando la notación de 46 con el subíndice* θ *puesto para enfatizar la dependencia en el parámetro, que*

$$\begin{aligned}h(\theta) :&= \frac{1}{n}\sum_{i=1}^n \left\langle \widehat{b}[\theta]-\mathtt{E}_\varepsilon(\widehat{b}[\theta])\,,\, x_i-\overline{x}\right\rangle(\varepsilon_i-\bar{\varepsilon}),\\ &= \sum_{i=1}^n n^{-1/2}(\varepsilon_i-\bar{\varepsilon})\left\langle \widehat{b}[\theta]-\mathtt{E}_\varepsilon(\widehat{b}[\theta])\,,\, n^{-1/2}(x_i-\overline{x})\right\rangle,\\ &= \left\langle\widetilde{\varepsilon},\, X_\theta(X_\theta^tX_\theta+\rho I)^{-1}X_\theta^t\widetilde{\varepsilon}\right\rangle,\\ &= \widetilde{\varepsilon}^tX_\theta(X_\theta^tX_\theta+\rho I)^{-1}X_\theta^t\widetilde{\varepsilon},\\ &= Tr(X_\theta(X_\theta^tX_\theta+\rho I)^{-1}X_\theta^t\widetilde{\varepsilon}\widetilde{\varepsilon}^t).\end{aligned}$$

Si bien podemos mostrar que $\mathtt{E}(h(\theta))\le\sigma^2\frac{p}{n}$ *esto no es aún suficiente ya que no es posible controlar un supremo por el supremo sobre las esperanza.*

Para obtener los resultados enunciados hacemos una expansión de Taylor de orden 1. Luego para θ_c *un punto en intermedio entre* θ *y* θ_0

$$h(\theta) = h(\theta_0) + dh(\theta_c)(\theta-\theta_0).$$

Para calcular la derivada utilizamos propiedades elementales de derivadas matriciales. Estas se pueden consultar en Minka (2000). Si llamamos $Z_\theta = (X_\theta^tX_\theta+\rho I)^{-1}$ *tenemos que*

$$dh(\theta) = d\left(Tr(X_\theta Z_\theta X_\theta^t\widetilde{\varepsilon}\widetilde{\varepsilon}^t)\right) = Tr\left(d(X_\theta Z_\theta X_\theta^t)\widetilde{\varepsilon}\widetilde{\varepsilon}^t\right).$$

Como $d(X_\theta Z_\theta X_\theta)$ *es simétrica y* $\widetilde{\varepsilon}\widetilde{\varepsilon}^t$ *semidefinida positiva tenemos por la Propiedad 11.2*

$$|dh(\theta)|\le\|d(X_\theta Z_\theta X_\theta^t)\|_2\|\widetilde{\varepsilon}\|^2.$$

Para derivar aplicamos la regla del producto y del inverso multiplicativo tenemos

$$d(X_\theta Z_\theta X_\theta^t) = d(X_\theta)Z_\theta X_\theta^t + X_\theta\left(d(Z_\theta)X_\theta^t + Z_\theta d(X_\theta^t)\right).$$

Como

$$d(Z_\theta) = -Z_\theta \left(d(X_\theta^t) X_\theta + X_\theta^t d(X_\theta) \right) Z_\theta \quad y \quad d(X_\theta^t) = d(X_\theta)^t.$$

tenemos que

$$d(X_\theta Z_\theta X_\theta^t) = (I - X_\theta Z_\theta X_\theta^t) d(X_\theta) Z_\theta X_\theta^t + X_\theta Z_\theta d(X_\theta)^t (I - X_\theta Z_\theta X_\theta^t). \tag{109}$$

Utilizando la desigualdad triangular, que los términos de la suma son sus respectivos transpuestos, que la norma 2 es submultiplicativa y la Propiedad (I.12) tenemos

$$\begin{aligned}
\|d(X_\theta Z_\theta X_\theta^t)\|_2 &\leq 2\|d(X_\theta) Z_\theta X_\theta^t\|_2 + 2\|X_\theta Z_\theta X_\theta^t d(X_\theta) Z_\theta X_\theta^t\|_2, \\
&\leq 2\|d(X_\theta) Z_\theta X_\theta^t\|_2 + 2\|X_\theta Z_\theta X_\theta^t\|_2 \|d(X_\theta) Z_\theta X_\theta^t\|_2, \\
&\leq 4\|d(X_\theta) Z_\theta X_\theta^t\|_2, \\
&\leq 4\|d(X_\theta)\|_2 \|Z_\theta\|_2 \|X_\theta\|_2, \\
&\leq 4\rho^{-1} \|d(X_\theta)\|_2 \|X_\theta\|_2.
\end{aligned}$$

Como las covariables tienen derivada continua, por la regla de Leibniz para derivar dentro de la integral, tenemos

$$\sqrt{n}\, d(X_\theta)_{ij} = \frac{3}{2}\theta^{1/2} \int_0^1 g_j(t)(x_i(\theta t) - \overline{x}(\theta t))\, dt + \theta^{1/2} \int_0^1 g_j(t)\theta t (x_i'(\theta t) - \overline{x}'(\theta t))\, dt.$$

Si definimos $\widetilde{X}_\theta \in \mathbb{R}^{n \times p}$ *como*

$$\begin{aligned}
(\widetilde{X}_\theta)_{ij} &= n^{-1/2} \int_0^\theta f_j(s) s (x'(s) - \overline{x}'(s))\, ds, \\
&= n^{-1/2} \theta^{3/2} \int_0^1 g_j(t) \theta t (x_i'(\theta t) - \overline{x}'(\theta t))\, dt,
\end{aligned}$$

resulta que $d(X_\theta) = \frac{1}{\theta}\left(\frac{3}{2} X_\theta + \widetilde{X}_\theta \right)$ *y por lo tanto*

$$\|d(X_\theta Z_\theta X_\theta^t)\|_2 \leq \frac{4}{\rho\theta} \left(\frac{3}{2} \|X_\theta\|_2^2 + \|X_\theta\|_2 \|\widetilde{X}_\theta\|_2 \right).$$

Veamos que lo que está dentro del paréntesis es acotado en probabilidad. Como la norma Frobenius es mayor o igual a la norma 2 y la base es ortonormal tenemos

$$\|X_\theta\|_2^2 \leq \|X_\theta\|_F^2 = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p \langle f_j\, ,\, x_i - \overline{x} \rangle^2 = \frac{1}{n} \sum_{i=1}^n \|\Pi_{H_\theta}(x_i - \overline{x})\|^2 \leq \frac{1}{n} \sum_{i=1}^n \|x_i - \overline{x}\|^2.$$

Si definimos la función $\tilde{w}_i(t) = t(x_i(t) - \overline{x}'(t))$

$$\|\widetilde{X}_\theta\|_2^2 \leq \|\widetilde{X}_\theta\|_F^2 = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p \langle f_j\, ,\, \tilde{w}_i \rangle^2 = \frac{1}{n} \sum_{i=1}^n \|\Pi_{H_\theta}(\tilde{w}_i)\|^2 \leq \frac{1}{n} \sum_{i=1}^n \|\tilde{w}_i\|^2.$$

Como $\frac{1}{\theta_c} \leq \frac{4}{\theta_0}$ *definimos*

$$B_n = \frac{16}{\theta_0}\|\widetilde{\varepsilon}\|^2 \left(\frac{3}{2}\frac{1}{n}\sum_{i=1}^{n}\|x_i - \overline{x}\|^2 + \frac{1}{n}\sqrt{\sum_{i=1}^{n}\|x_i - \overline{x}\|^2 \sum_{i=1}^{n}\|\tilde{w}_j\|^2} \right),$$

por lo tanto $B_n = \mathcal{O}_P(1)$. *Si definimos*

$$Y_{H,\delta,\rho,n} = h(\theta_0) + \frac{4}{n}\sum_{i=1}^{n}\|x_i - \overline{x}\|^2 \left(\sup_{\theta\in[\theta_0-\delta,\theta_0+\delta]} \|\Pi_{H_\theta}(b_\theta) - b_\theta\|^2 + \delta\|b\|^2 \right) + \rho\|b\|^2.$$

Tenemos que $h(\theta_0) \geq 0$ *y en consecuencia* $Y_{H,\delta,\rho,n} \geq 0$. *Además como* $\mathtt{E}(h(\theta_0)) \leq \sigma^2\frac{p}{n}$ *tomando*

$$C = \max\left\{\sigma^2, \|b\|^2, 4\mathtt{E}(\|x-m\|^2)(1+\|b\|^2)\right\},$$

tenemos

$$E(Y_{H,\delta}) \leq C\left(\frac{p}{n} + \sup_{\theta\in[\theta_0-\delta,\theta_0+\delta]} \|\Pi_{H_\theta}(b_\theta) - b_\theta\|^2 + \rho + \delta \right).$$

Finalmente hemos probado la descomposición enunciada ya que

$$\sup_{\theta\in[\theta_0-\delta,\theta_0+\delta]} \|\widehat{b}[\theta] - b\|^2_{\Gamma_n} + \rho\|\widehat{b}[\theta]\|^2 \leq Y_{H,\delta,\rho,n} + \frac{\delta}{\rho}B_n,$$

10.4 (Proposición 4.7) *Queremos ver que dado* $\varepsilon > 0$ *existe* $K > 0$ *y* n_0 *tal que si* $n \geq n_0$

$$P\left(\|\widehat{b}_n[\hat{\theta}_n] - b\|^2_{\Gamma_n} > K\mathtt{d}_n^{1/2} \right) < \varepsilon.$$

Definimos los eventos y las variables

$$A_n = \{|\hat{\theta}_n - \theta_0| > K_1\mathtt{d}_n\},$$
$$G_n = \|\widehat{b}_n[\hat{\theta}_n] - b\|^2_{\Gamma_n}\mathbb{I}_{A_n}.$$

donde K_1 *es tal que existe* n_1 *para el cual* $P(A_n) \leq \varepsilon/3$ *si* $n \geq n_1$. *Si llamamos* $\delta_n = K_1\mathtt{d}_n$ *obtenemos la siguiente desigualdad*

$$\begin{aligned}\|\widehat{b}_n[\hat{\theta}_n] - b\|^2_{\Gamma_n} &\leq \sup_{\theta\in[\theta_0-\delta_n,\theta_0+\delta_n]} \|\widehat{b}_n[\theta] - b\|^2_{\Gamma_n} + G_n \\ &\leq \sup_{\theta\in[\theta_0-\delta_n,\theta_0+\delta_n]} \|\widehat{b}_n[\theta] - b\|^2_{\Gamma_n} + \rho_n\|\widehat{b}[\theta]\|^2 + G_n\end{aligned}$$

Por el Teorema 4.6 existen $Y_{H_n,\delta_n,\rho_n,n}$ *y* B_n *sucesiones de variables no negativas tales que*

$$\|\widehat{b}_n[\hat{\theta}_n] - b\|^2_{\Gamma_n} \leq Y_{H_n,\rho_n,\delta_n,n} + \frac{\delta_n}{\rho_n}B_n + G_n.$$

Por lo tanto tenemos que para $K > 0$

$$P\left(\|\widehat{b}_n[\hat{\theta}_n] - b\|^2_{\Gamma_n} > 3K\mathtt{d}_n^{1/2}\right) \leq P(Y_{H_n,\rho_n,\delta_n,n} \geq K\mathtt{d}_n^{1/2}) + P(B_n \geq K\rho_n\delta_n^{-1}\mathtt{d}_n^{1/2}) + P(G_n \geq K\mathtt{d}_n^{1/2}).$$

Resta ver que existe un n_0 *tal que para todo* $n \geq n_0$*, cada uno de los sumandos es menor o igual a* $\varepsilon/3$ *si* K *es suficientemente grande.*

Notemos que si $n \geq n_1$

$$P\left(G_n \geq K\mathtt{d}_n^{1/2}\right) \leq P(A_n) \leq \varepsilon/3.$$

Por el Teorema 4.6 y la Desigualdad de Markov resulta que $Y_{H_n,\delta_n,\rho_n,n} = \mathcal{O}_P\left(\mathtt{d}_n^{1/2}\right)$*. Por lo que si* K *es suficientemente grande, existe* n_2 *tal que si* $n \geq n_2$

$$P\left(Y_{H_n,\delta_n,\rho_n,n} \geq K\mathtt{d}_n^{1/2}\right) \leq \varepsilon/3.$$

Como $\rho_n\delta_n^{-1}\mathtt{d}_n^{1/2} \propto 1$ *y* $B_n = \mathcal{O}_P(1)$ *tenemos que si* K *es lo suficientemente grande existe* n_3 *tal que si* $n \geq n_3$

$$P\left(B_n \geq K\rho_n\delta_n^{-1}\mathtt{d}_n^{1/2}\right) \leq \varepsilon/3.$$

Para probar que la norma cuadrada del estimador está acotada en probabilidad definimos las variables aleatorias

$$F_n = \|\widehat{b}_n[\hat{\theta}_n]\|^2\mathbb{I}_{A_n}.$$

Del Teorema 4.6 tenemos que

$$\begin{aligned}\rho_n\|\widehat{b}_n[\hat{\theta}_n]\|^2 &\leq \sup_{\theta\in[\theta_0-\delta_n,\theta_0+\delta_n]} \rho_n\|\widehat{b}_n[\theta]\|^2 + G_n,\\ &\leq Y_{H_n,\rho_n,\delta_n,n} + \frac{\delta_n}{\rho_n}B_n + G_n.\end{aligned}$$

Como $P(G_n \geq K\mathtt{d}_n^{1/2}) \leq P(A_n) \leq \varepsilon/3$ *obtenemos*

$$\rho_n\|\widehat{b}_n[\hat{\theta}_n]\|^2 = \mathcal{O}_P\left(\mathtt{d}_n^{1/2}\right).$$

Como $\rho_n \propto \mathtt{d}_n^{1/2}$ *resulta*

$$\|\widehat{b}_n[\hat{\theta}_n]\|^2 = \mathcal{O}_P(1).$$

10.5 (Corolario 4.8) *Por el Teorema 3.13 tenemos*

$$\|\widetilde{b}_n[\hat{\theta}_n] - b\|^2_{\Gamma_n} \leq 4\|\widehat{b}_n[\hat{\theta}_n] - b\|^2_{\Gamma_n} + 6\|b - \Pi_{E_{n\hat{\theta}_n}}(b)\|^2_{\Gamma_n} + 8\rho_n\|\widehat{b}[\hat{\theta}_n] - \Pi_{E_{n\hat{\theta}_n}}(b)\|^2$$

$$\frac{1}{2}\|\widetilde{b}_n[\hat{\theta}_n]\|^2 \leq \|\widehat{b}_n[\hat{\theta}_n]\|^2 + \frac{2}{\rho_n}\left(\|\widehat{b}_n[\hat{\theta}_n] - b\|^2_{\Gamma_n} + \|b - \Pi_{E_{n\hat{\theta}_n}}(b)\|^2_{\Gamma_n}\right) + \|\widehat{b}_n[\hat{\theta}_n] - \Pi_{E_{n\hat{\theta}_n}}(b)\|^2$$

Observando que $\Pi_{E_{n,\theta}}(b) = \Pi_{E_{n,\theta}}(b_\theta)$ *tenemos*

$$\|\widehat{b}[\hat{\theta}_n] - \Pi_{E_{n\hat{\theta}_n}}(b)\|^2 \leq 2\|\widehat{b}[\hat{\theta}_n] - b\|^2 + 2\|b - \Pi_{E_{n\hat{\theta}_n}}(b)\|^2 = \mathcal{O}_P(1),$$
$$\|b - \Pi_{E_{n\hat{\theta}_n}}(b)\|^2 \leq \|b_{\hat{\theta}_n} - \Pi_{E_{n\hat{\theta}_n}}(b_{\hat{\theta}_n})\|^2 + \|b\|_\infty^2 |\hat{\theta}_n - \theta_0| = \mathcal{O}_P(\rho_n).$$

Como $\frac{1}{n}\sum_{i=1}^n \|x_i - \overline{x}\|^2 = \mathcal{O}_P(1)$ *utilizando el Lema 8.13 apropiadamente obtenemos*

$$\|\widetilde{b}_n[\hat{\theta}_n] - b\|_{\Gamma_n}^2 = \mathcal{O}_P(\rho_n) \qquad y \qquad \|\widetilde{b}_n[\hat{\theta}_n]\|^2 = \mathcal{O}_P(1).$$

11. Apendice

11.1. Propiedades de la traza

A continuación listamos algunas propiedades de la traza de matrices.

(I.1) La esperanza de una forma cuadrática de un vector aleatorio ε con $\mu = \mathtt{E}(\varepsilon)$ y $\Sigma = \mathtt{Var}(\varepsilon)$

$$\mathtt{E}(\varepsilon^t A\varepsilon) = \mathrm{Tr}(A\Sigma) + \mu^t A\mu.$$

(I.2) La traza es invariante ante permutaciones cíclicas, es decir

$$\mathrm{Tr}(ABC) = \mathrm{Tr}(BCA) = \mathrm{Tr}(CAB).$$

(I.3) Si $0 \leq A \leq B$. La desigualdad se mantiene por multiplicaciones a ambos lados para cualquier matriz C

$$0 \leq C^t AC \leq C^t BC.$$

(I.4) Si A es una matriz semidefinida positiva entonces existe una raíz cuadrada simétrica $A^{1/2}$ tal que $A^{1/2}A^{1/2} = A$.

(I.5) Si A y B son matrices definidas positivas $0 < B < A$, entonces

$$0 < A^{-1} < B^{-1}.$$

(I.6) Si $0 \leq B \leq C$ y $0 \leq A$ entonces $\mathrm{Tr}(AB) \leq \mathrm{Tr}(AC)$.

(I.7) Si $0 \leq B \leq A$ y A es inversible, entonces

$$\mathrm{Tr}(\left(A^{-1}B\right)^2) \leq \mathrm{Tr}\left(A^{-1}B\right).$$

La desigualdad es estricta si $A > B$.

(I.8) Si $A > 0$ y $C > 0$ y $B = B_1 + B_2$ con $B_i \geq 0$ para $i = 1, 2$, entonces

$$\mathrm{Tr}\left((A+B)^{-1}B\right) \leq \mathrm{Tr}((A+B_1)^{-1}B_1) + \mathrm{Tr}((A+B_2)^{-1}B_2)$$

(I.9) Sean A B matrices semidefinidas positivas. Entonces $\mathrm{Tr}(AB) \leq \sum_{j=1}^{p} \tau_j(A)\tau_j(B)$ donde $\tau_j(A)$ y $\tau_j(B)$ son los j ésimos autovalores de A y B.

(I.10) Sean A y B matrices semidefinidas positivas. Notemos el k ésimo autovalor mas pequeño de una matriz por $\tau_k(X)$, entonces

$$\tau_k(A+B) \geq \mathrm{máx}\{\tau_k(A), \tau_k(B)\}.$$

(I.11) Para $A \geq 0$ y $B = B_1 + B_2$ con $B_i \geq 0$ para $i = 1, 2$. Si $\dagger$ denota la inversa de Moore-Penrose de una matriz, resulta

$$\mathrm{Tr}((A+B)^\dagger B) \leq \mathrm{Tr}((A+B_1)^\dagger B_1) + \mathrm{Tr}((A+B_2)^\dagger B_2).$$

Dem: Si X, Y son matrices semidefinidas positivas y $P = XX^\dagger$ denota la proyección ortogonal sobre el espacio generado por las columnas de X, entonces $P(X+Y)^\dagger P \preceq X^\dagger$. Esto se puede ver usando Complementos de Schur.

Si $X = A + B_1$ e $Y = B_2$ obtenemos

$$B_1^{1/2} P (A + B_1 + B_2)^\dagger P B_1^{1/2} \preceq B_1^{1/2} (A+B_1)^\dagger B_1^{1/2}.$$

El resultado se obtiene notando que $B_1^{1/2} P = P B_1^{1/2} = B_1^{1/2}$.

(I.12) Si $A \geq 0$ y $B > 0$ con $D = B^{-1/2} A B^{-1/2}$ vale que

$$\begin{aligned} \mathrm{Tr}((A+B)^{-1}A) &= \mathrm{Tr}((D+I)^{-1}D), \\ \tau_{max}((A+B)^{-1}A) &= \tau_{max}((D+I)^{-1}D)). \end{aligned}$$

Además $0 \leq (D+I)^{-1} D < Id$ y $(D+I)^{-1} < D$.

Dem: Las igualdades para la traza y el máximo autovalor se obtienen de aplicar la propiedad (I.2) o que el espectro del producto de matrices es invariante por permutaciones.

$$\begin{aligned} (A+B)^{-1}A &= \left(B^{1/2}(B^{-1/2}AB^{-1/2} + I)B^{1/2}\right)^{-1} A, \\ &= B^{-1/2}(B^{-1/2}AB^{-1/2} + I)^{-1} B^{-1/2} A. \end{aligned}$$

Como $(D+I)^{-1}D = I - (D+I)^{-1}$ resulta una matriz simétrica. Por la propiedad (I.5) resulta $0 < I - (D+I)^{-1} < I$. Además, si $Dv = \lambda v$ entonces $(D+I)^{-1} D v = \frac{\lambda}{1+\lambda} v$ y por lo tanto todos los autovalores estan entre 0 y 1.

Proposición 11.1 *Si $A \in \mathbb{R}^{n \times p}$ y $B \in \mathbb{R}^{p \times n}$ entonces el espectro de AB coincide con el de BA. Es decir tienen el mismo conjunto de autovalores.*

Proposición 11.2 *Si A es simétrica y B semidefinida positiva entonces*

$$\lambda_{min}(A)\, Tr(B) \leq Tr(AB) \leq \lambda_{max}(A)\, Tr(B).$$

Dem: Primero notemos que si A y B son ambas definidas positivas entonces

$$Tr(AB) = Tr(A^{1/2} B A^{1/2}) \geq 0.$$

Sabiendo esto consideramos $A - \lambda_{min}(A) I \geq 0$ y $\lambda_{max}(A) I - B \geq 0$. Por lo tanto

$$0 \geq Tr((A - \lambda_{min}(A))B) = Tr(AB) - \lambda_{min}(A)\, Tr(B).$$

La otra desigualdad se obtiene de manera análoga.

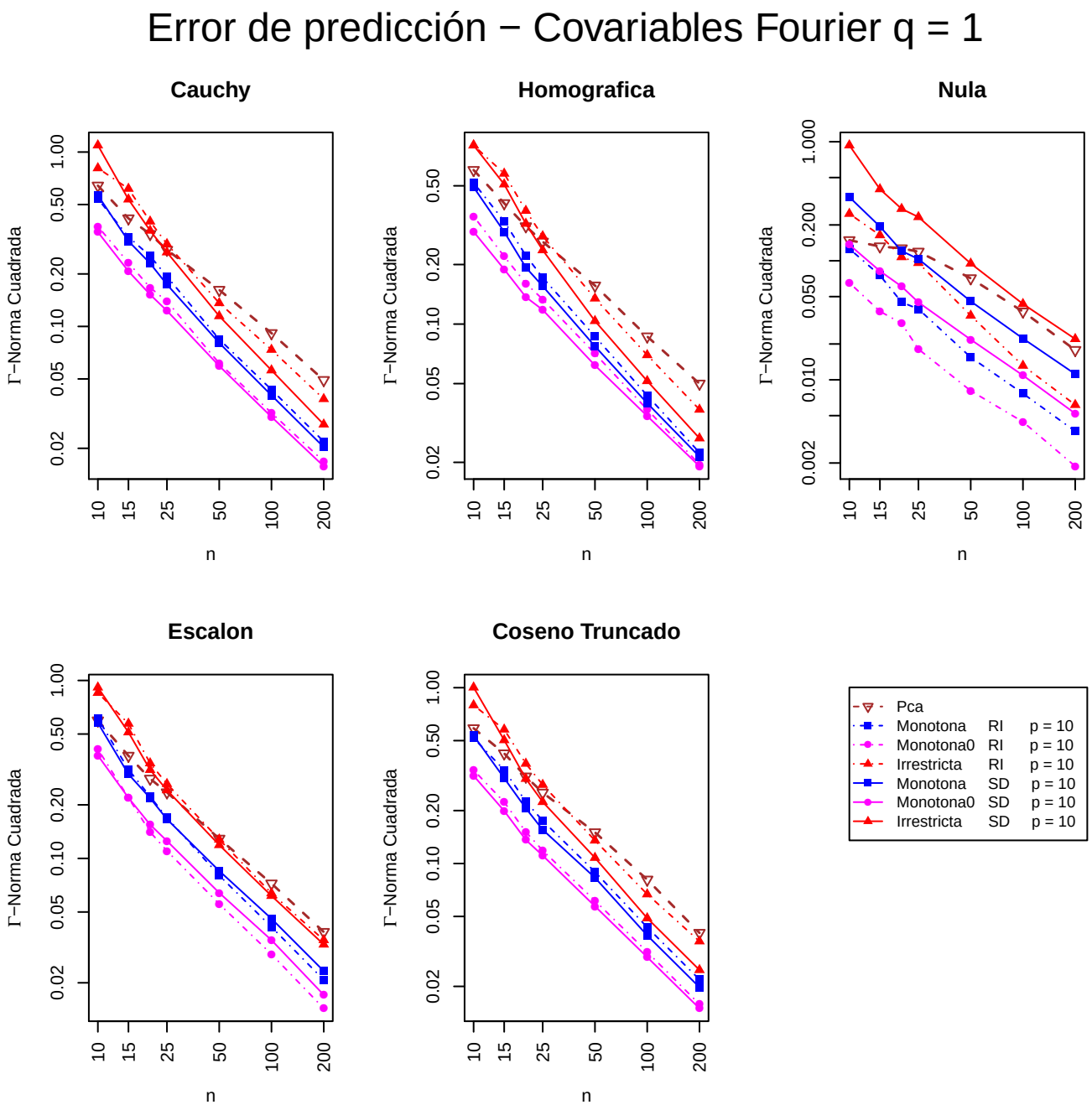

Figura 18: Curvas log-log de $\mathtt{E}\left(\|\widehat{b}_n - b\|^2_\Gamma\right)$ para estimadores con restricciones de monotonía y covariables Fourier $q = 1$.

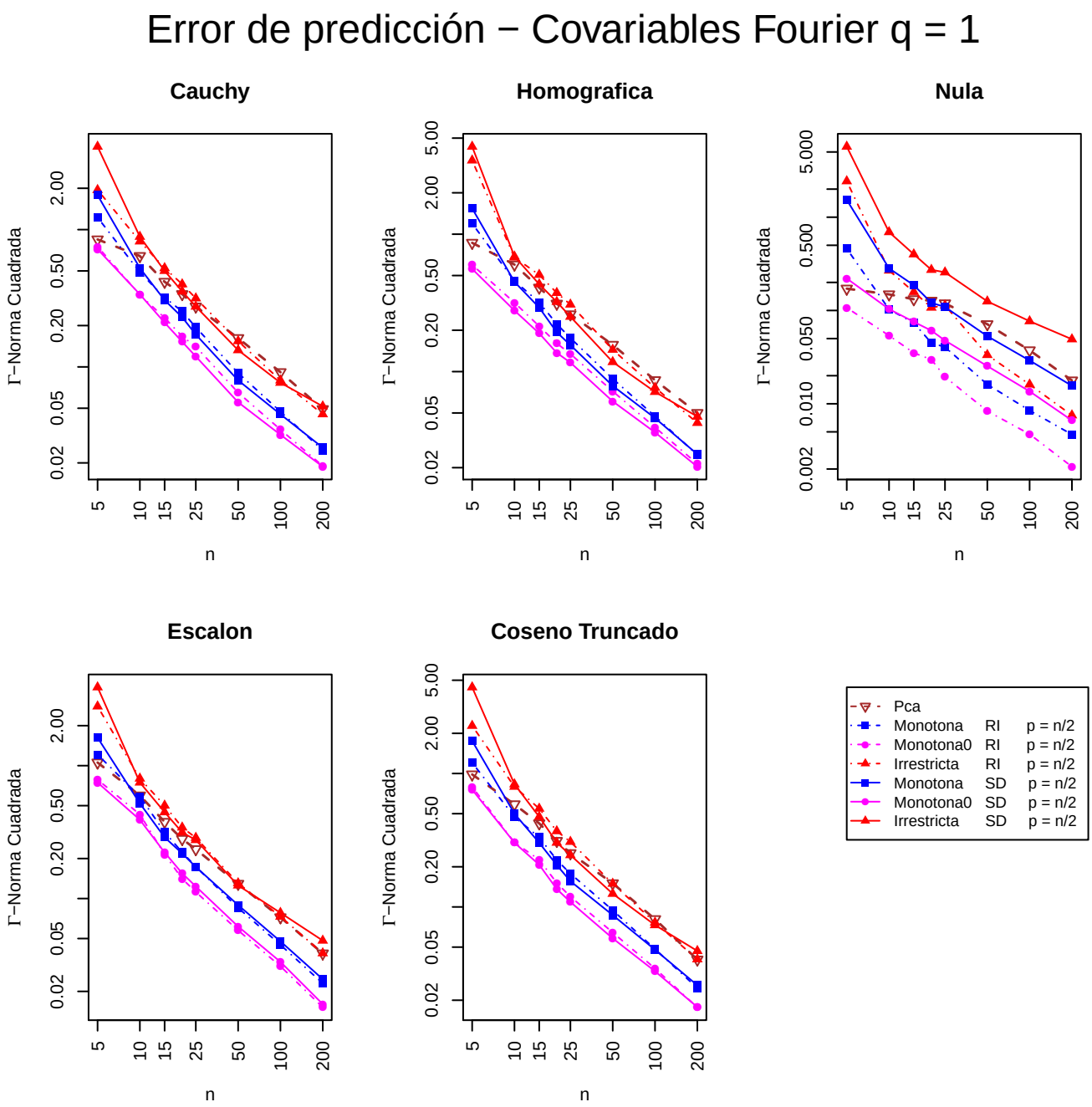

Figura 19: Curvas log-log de $\mathtt{E}\left(\|\widehat{b}_n - b\|_\Gamma^2\right)$ para estimadores con restricciones de monotonía y covariables Fourier $q = 1$.

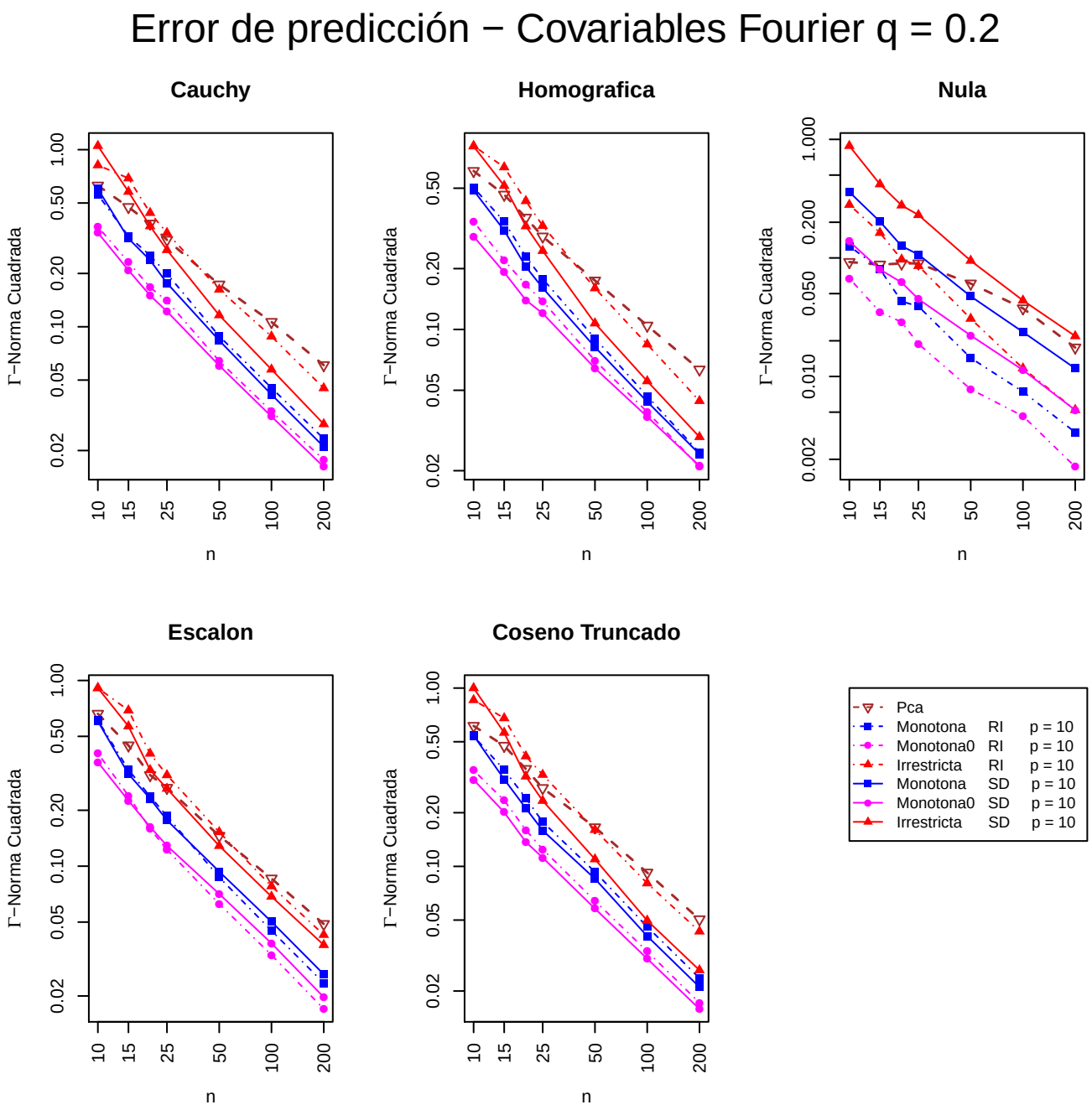

Figura 20: Curvas log-log de $\mathtt{E}\left(\|\widehat{b}_n - b\|^2_\Gamma\right)$ para estimadores con restricciones de monotonía y covariables Fourier $q = 0{,}2$.

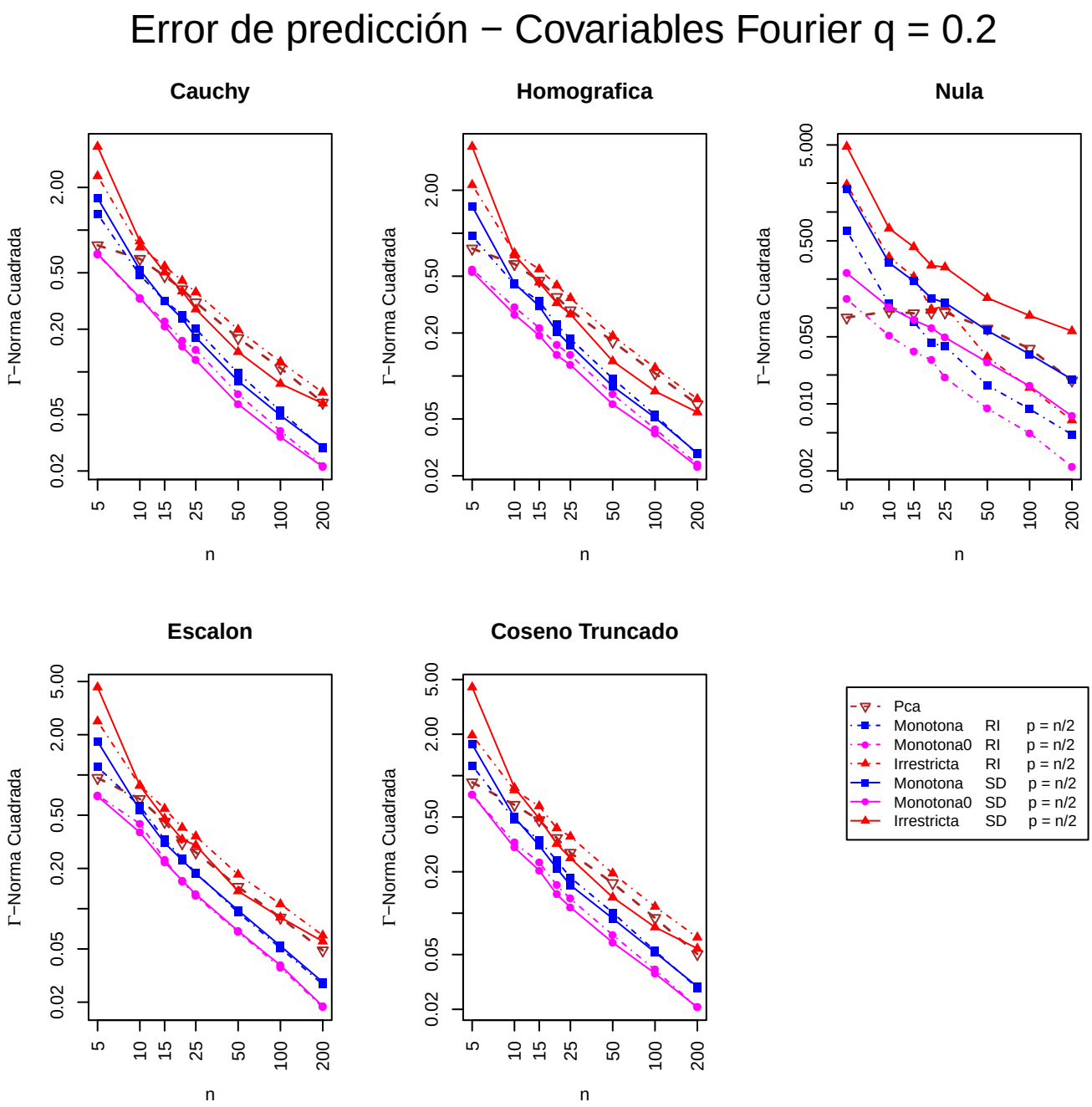

Figura 21: Curvas log-log de $\mathtt{E}\left(\|\widehat{b}_n - b\|_\Gamma^2\right)$ para estimadores con restricciones de monotonía y covariables Fourier $q = 0{,}2$.

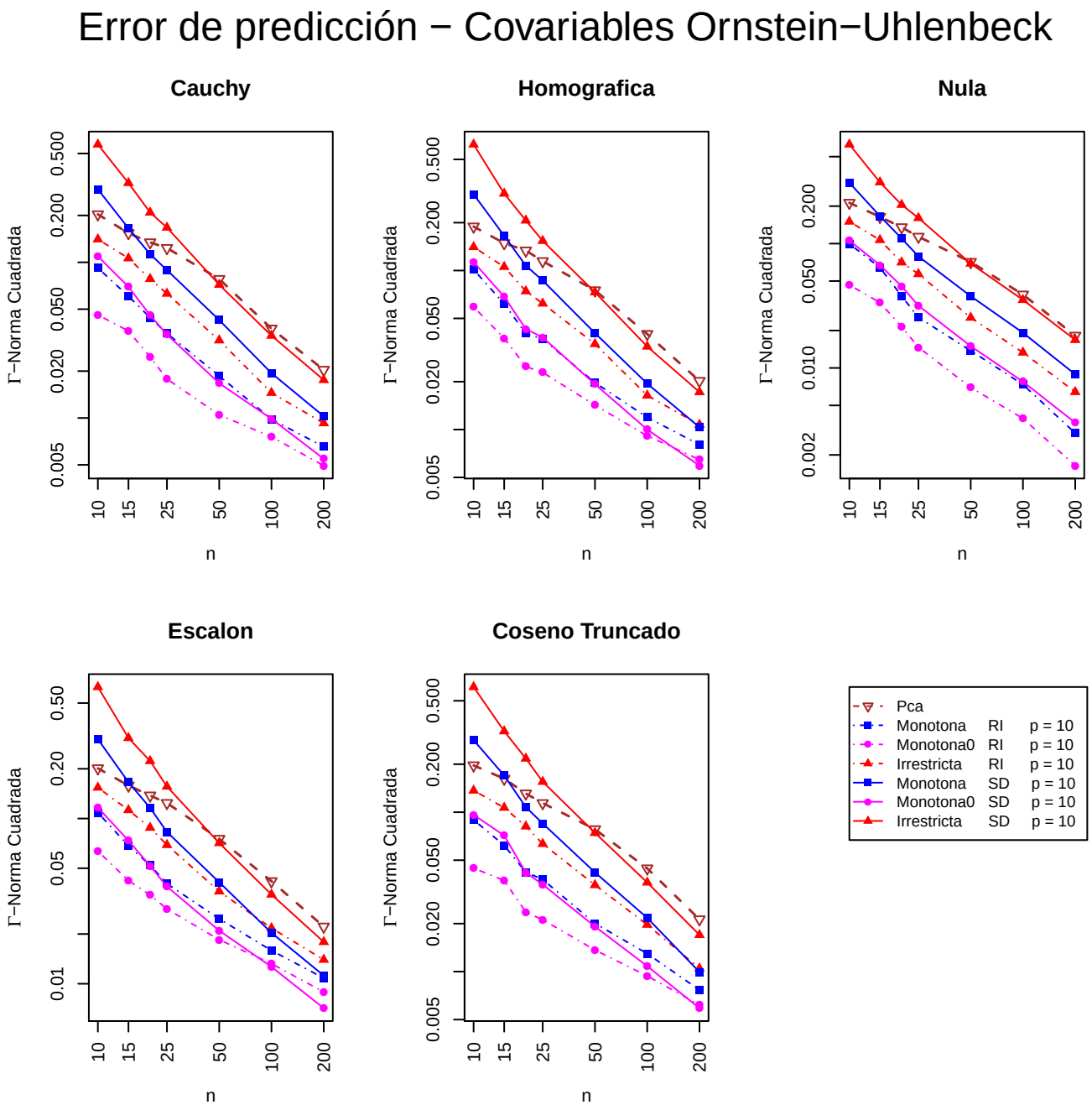

Figura 22: Curvas log-log de $\mathtt{E}\left(\|\widehat{b}_n - b\|_\Gamma^2\right)$ para estimadores con restricciones de monotonía y covariables Ornstein-Uhlenbeck.

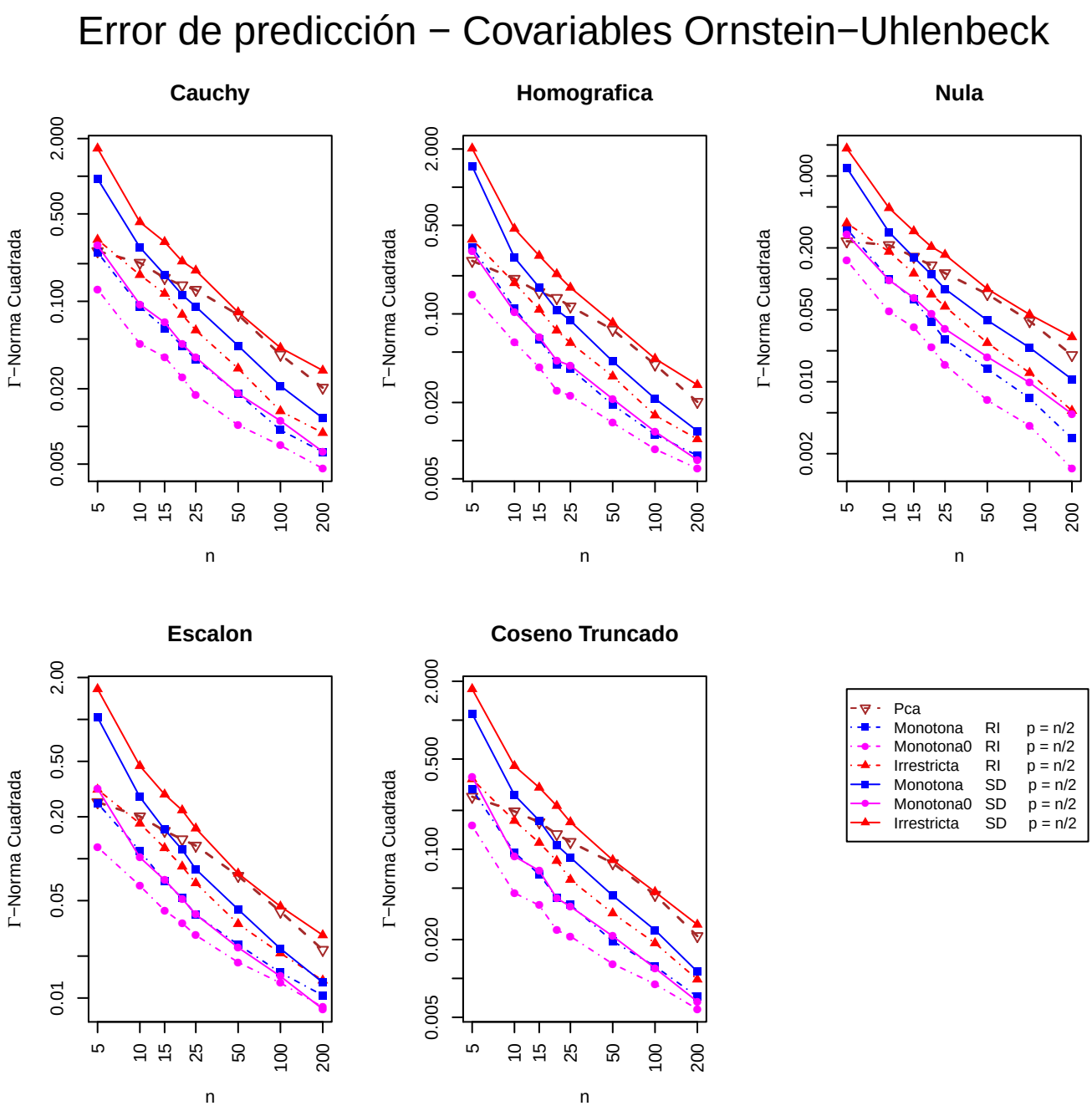

Figura 23: Curvas log-log de $\mathtt{E}\left(\|\widehat{b}_n - b\|_\Gamma^2\right)$ para estimadores con restricciones de monotonía y covariables Ornstein-Uhlenbeck.

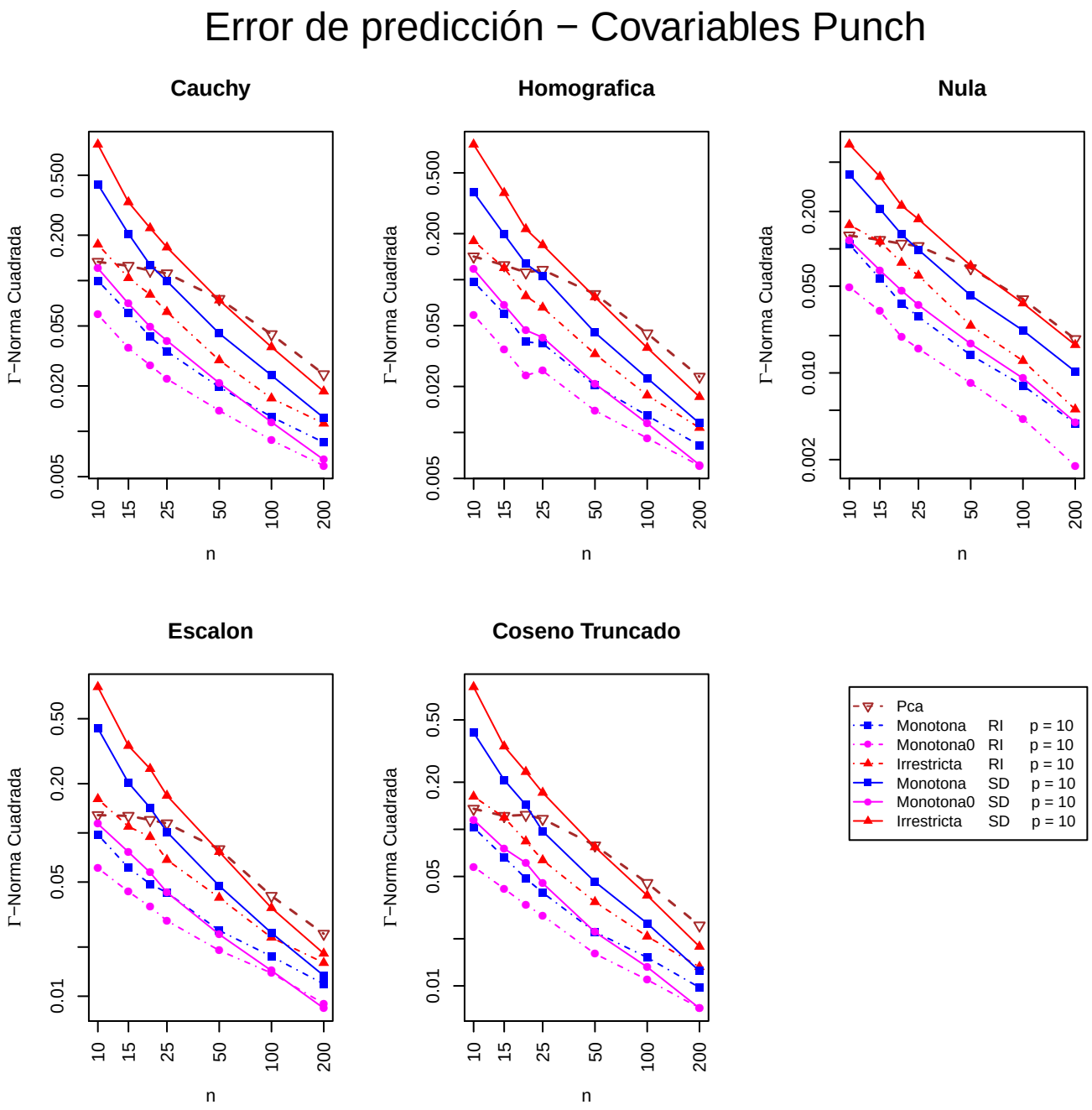

Figura 24: Curvas log-log de $\mathtt{E}\left(\|\widehat{b}_n - b\|_\Gamma^2\right)$ para estimadores con restricciones de monotonía y covariables Punch.

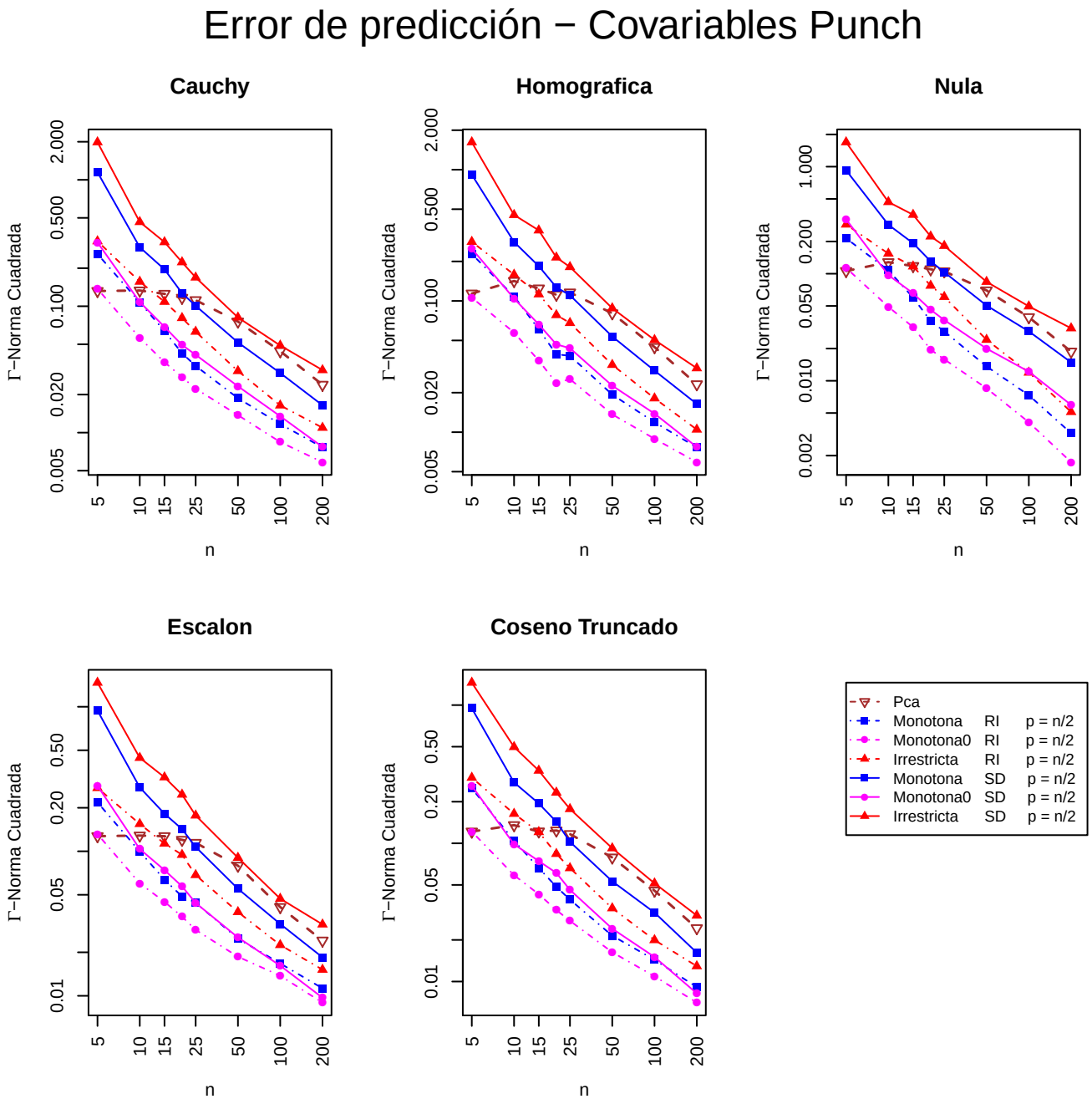

Figura 25: Curvas log-log de $\mathtt{E}\left(\|\widehat{b}_n - b\|_\Gamma^2\right)$ para estimadores con restricciones de monotonía y covariables Punch.

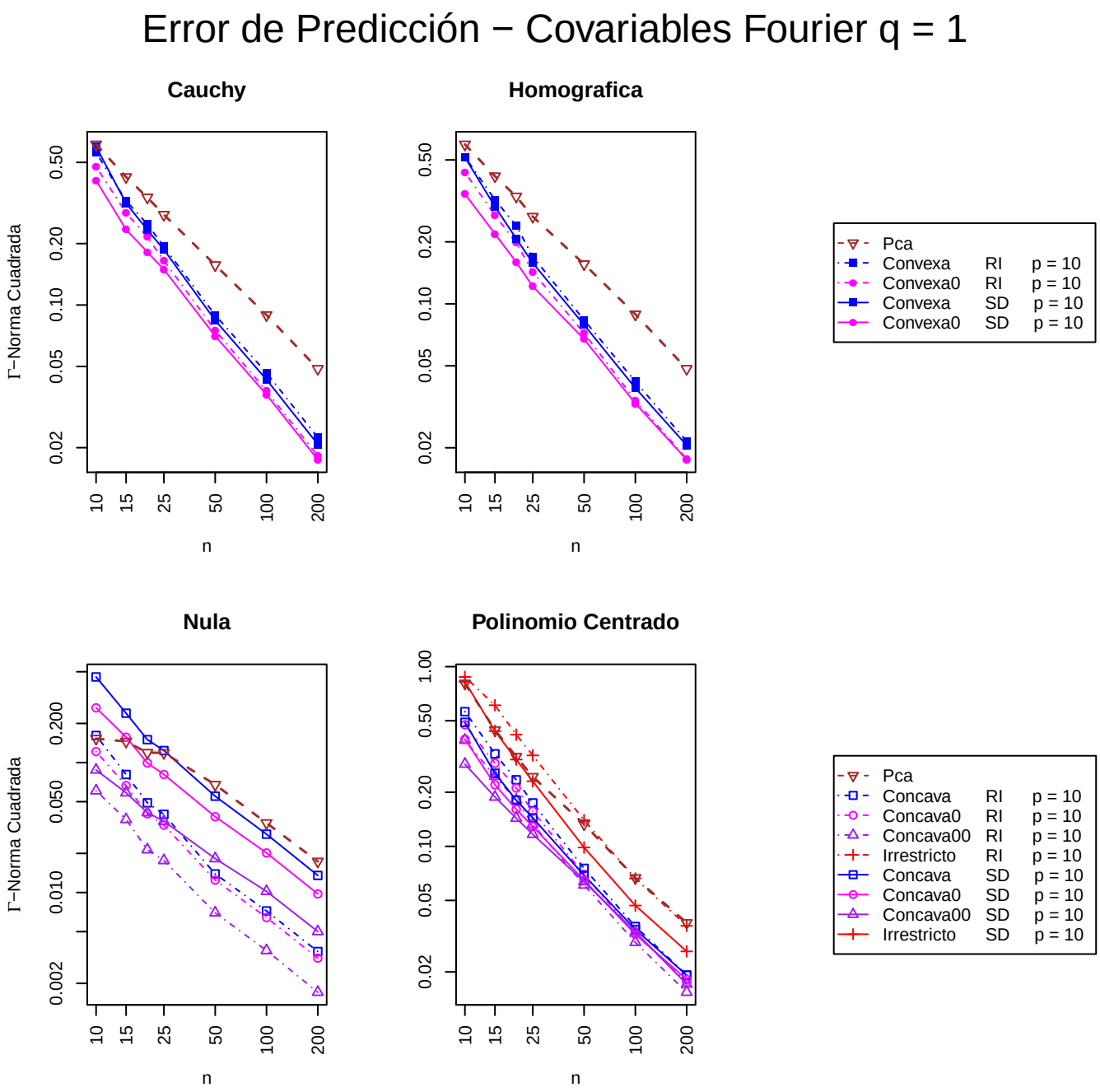

Figura 26: Curvas log-log de $\mathtt{E}\left(\|\widehat{b}_n - b\|^2_\Gamma\right)$ para estimadores con restricciones de concavidad y covariables Fourier $q = 1$.

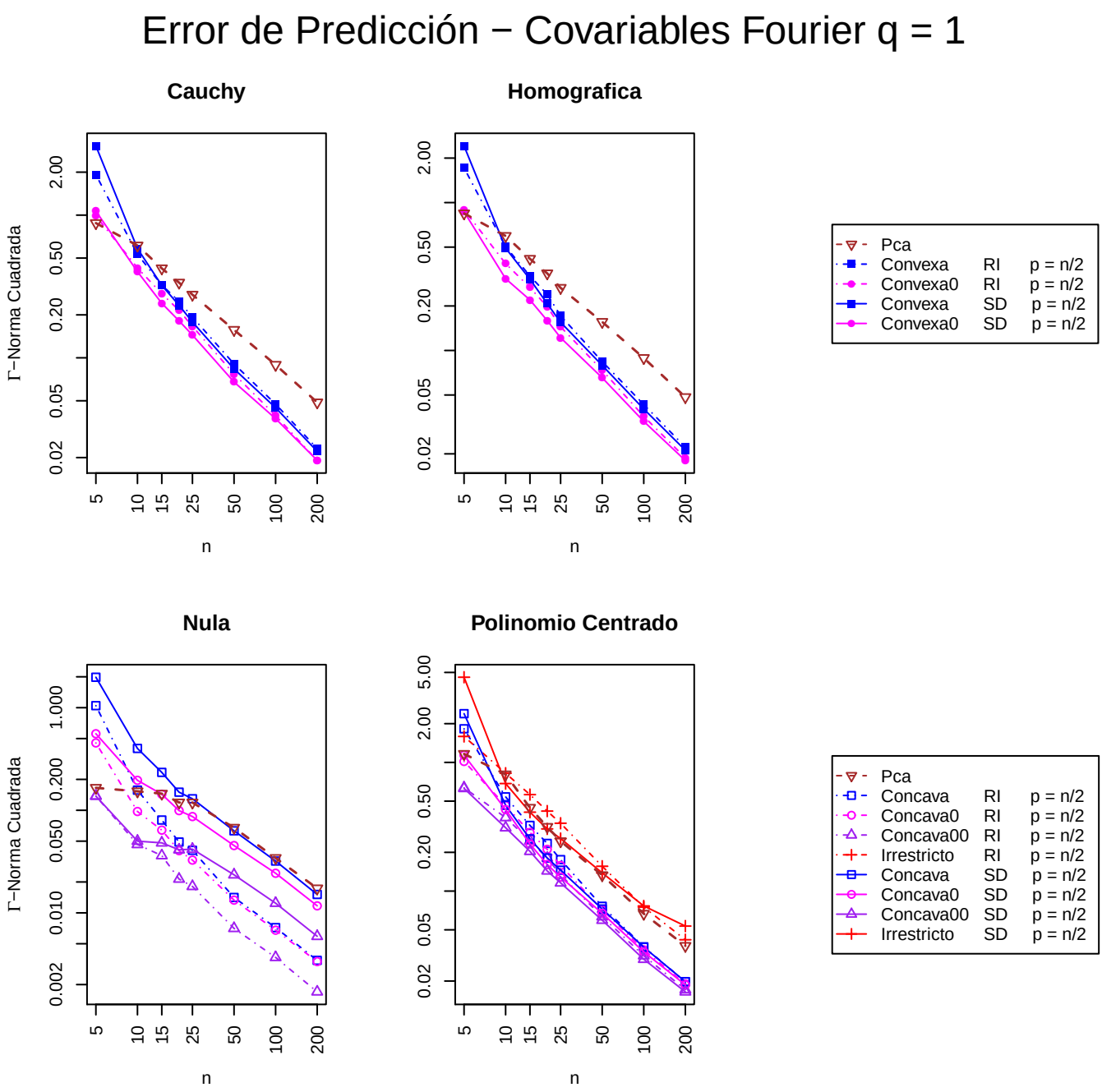

Figura 27: Curvas log-log de $\mathtt{E}\left(\|\widehat{b}_n - b\|^2_\Gamma\right)$ para estimadores con restricciones de concavidad y covariables Fourier $q = 1$.

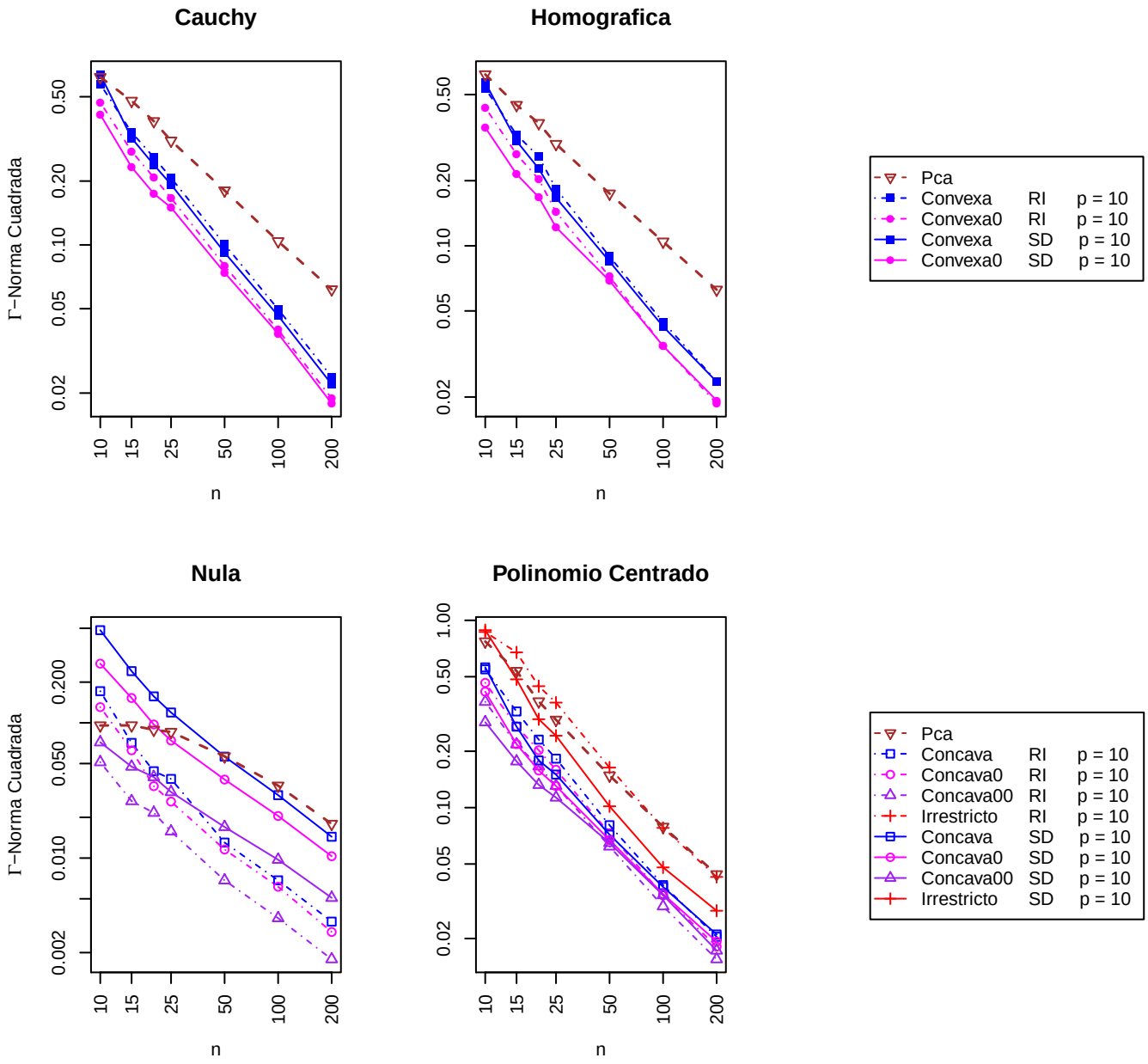

Figura 28: Curvas log-log de $\mathtt{E}\left(\|\widehat{b}_n - b\|^2_\Gamma\right)$ para estimadores con restricciones de concavidad y covariables Fourier $q = 0{,}2$.

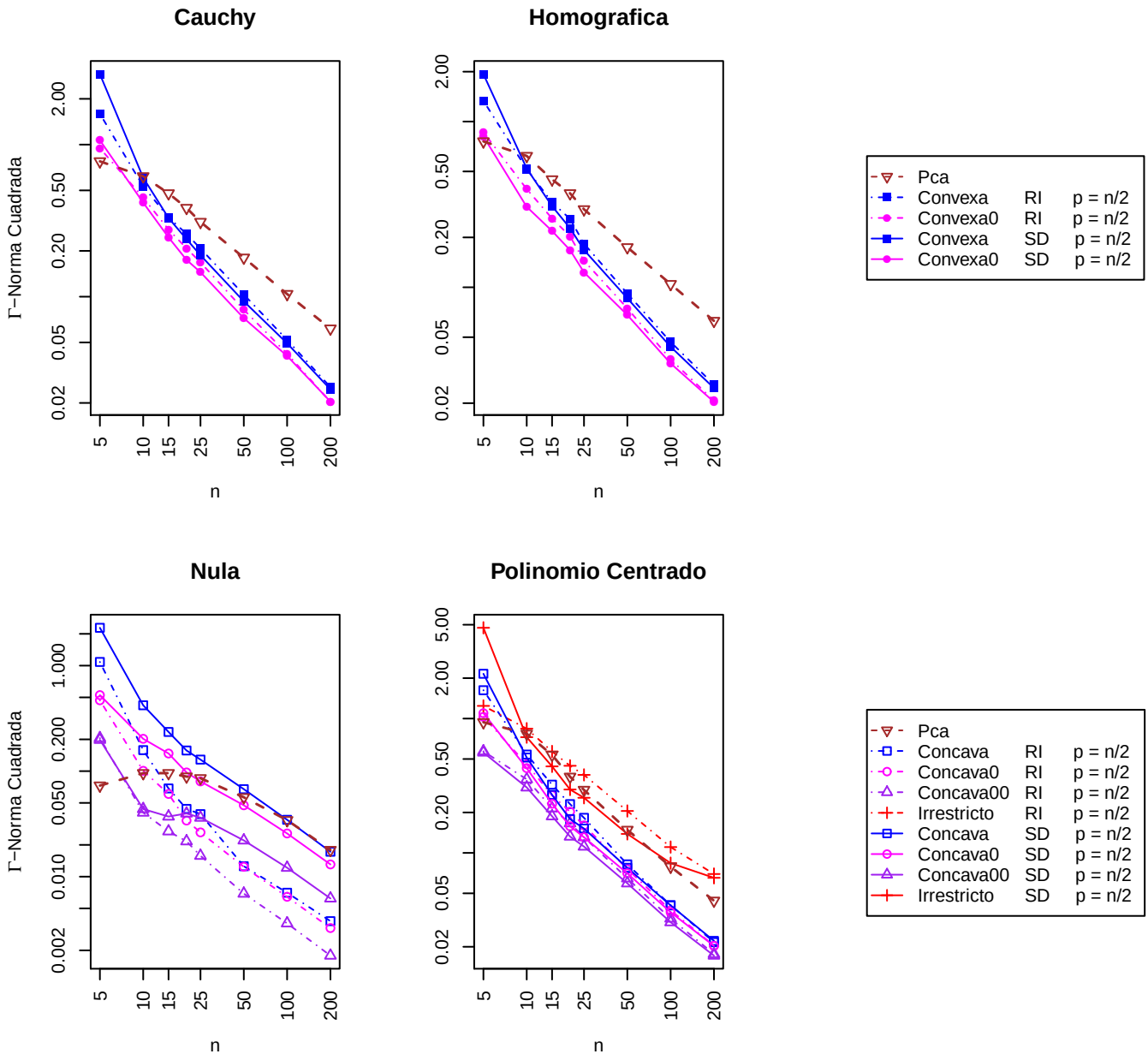

Figura 29: Curvas log-log de $\mathtt{E}\left(\|\widehat{b}_n - b\|_\Gamma^2\right)$ para estimadores con restricciones de concavidad y covariables Fourier $q = 0{,}2$.

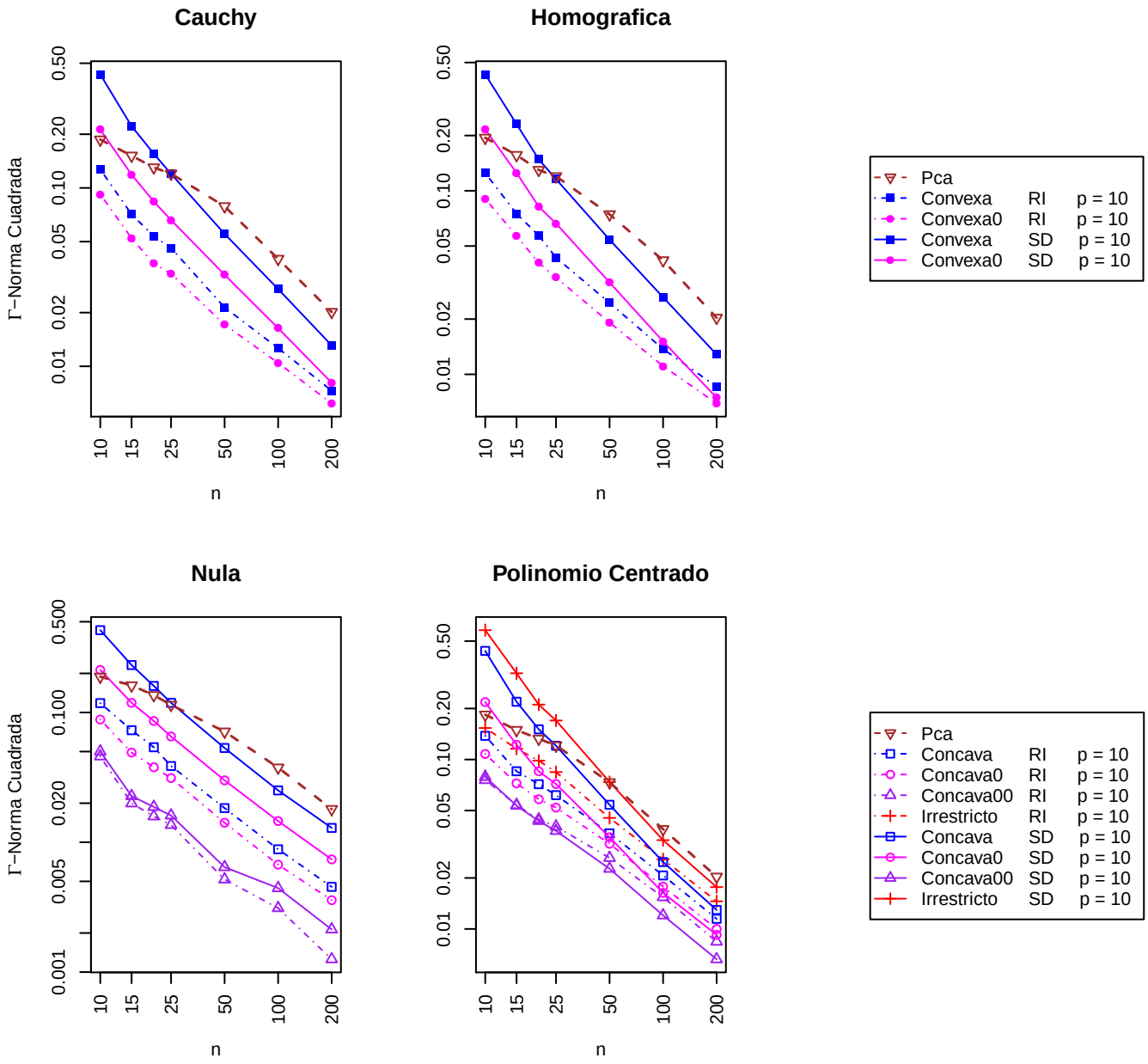

Figura 30: Curvas log-log de $\mathtt{E}\left(\|\widehat{b}_n - b\|^2_\Gamma\right)$ para estimadores con restricciones de concavidad y covariables Ornstein-Uhlenbeck.

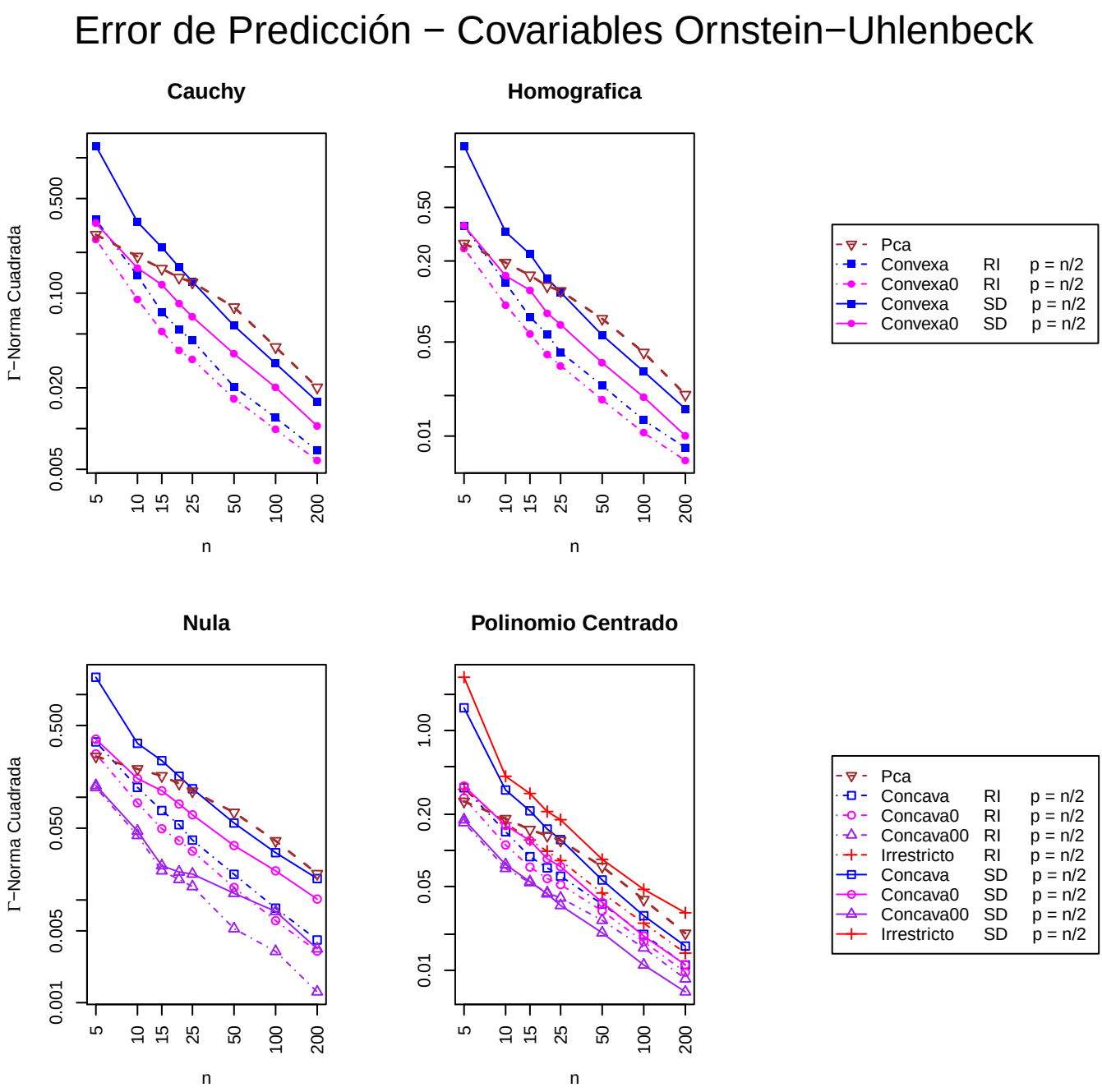

Figura 31: Curvas log-log de $\mathtt{E}\left(\|\widehat{b}_n - b\|_\Gamma^2\right)$ para estimadores con restricciones de concavidad y covariables Ornstein-Uhlenbeck.

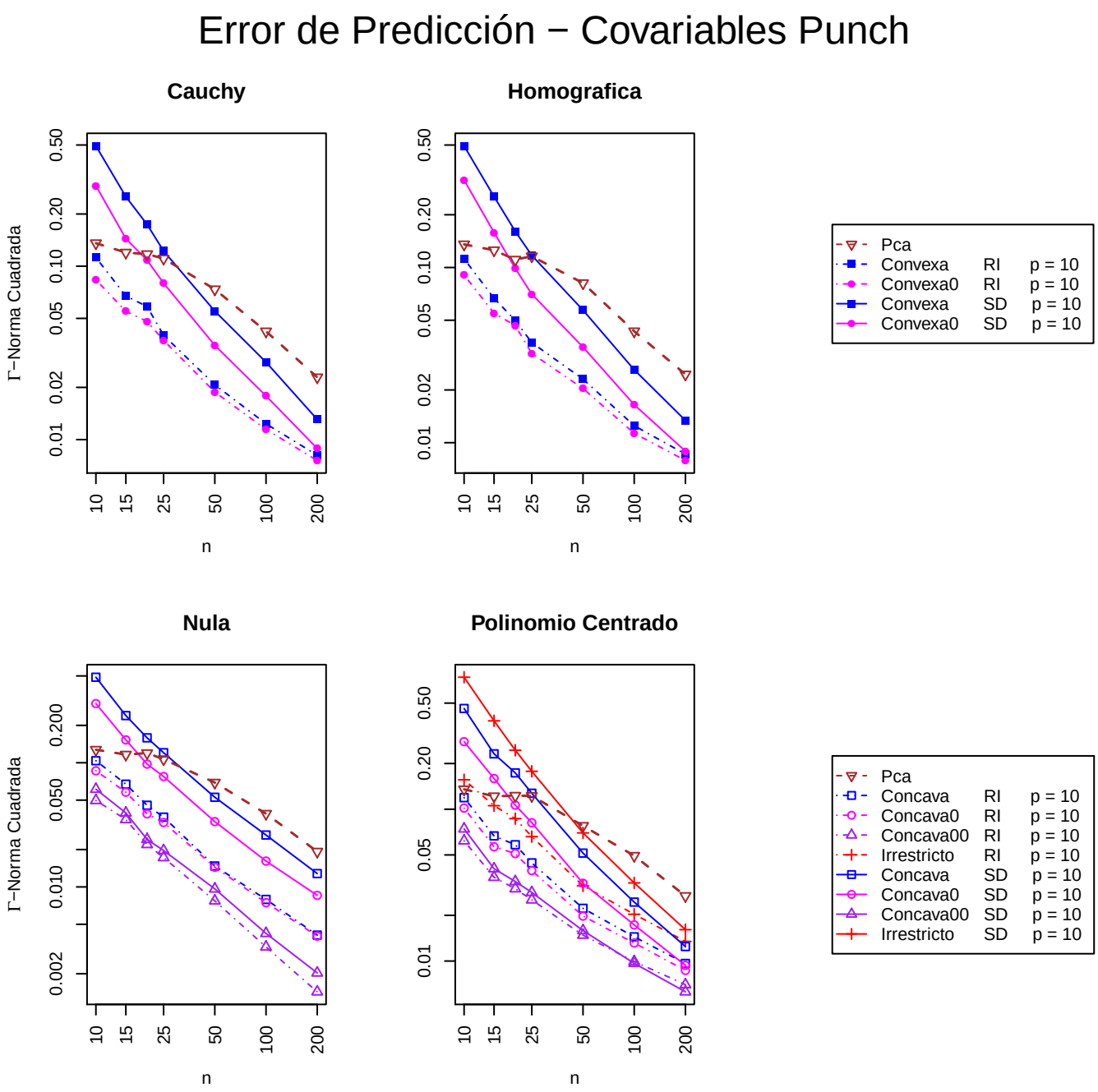

Figura 32: Curvas log-log de $\mathtt{E}\left(\|\widehat{b}_n - b\|^2_\Gamma\right)$ para estimadores con restricciones de concavidad y covariables Punch.

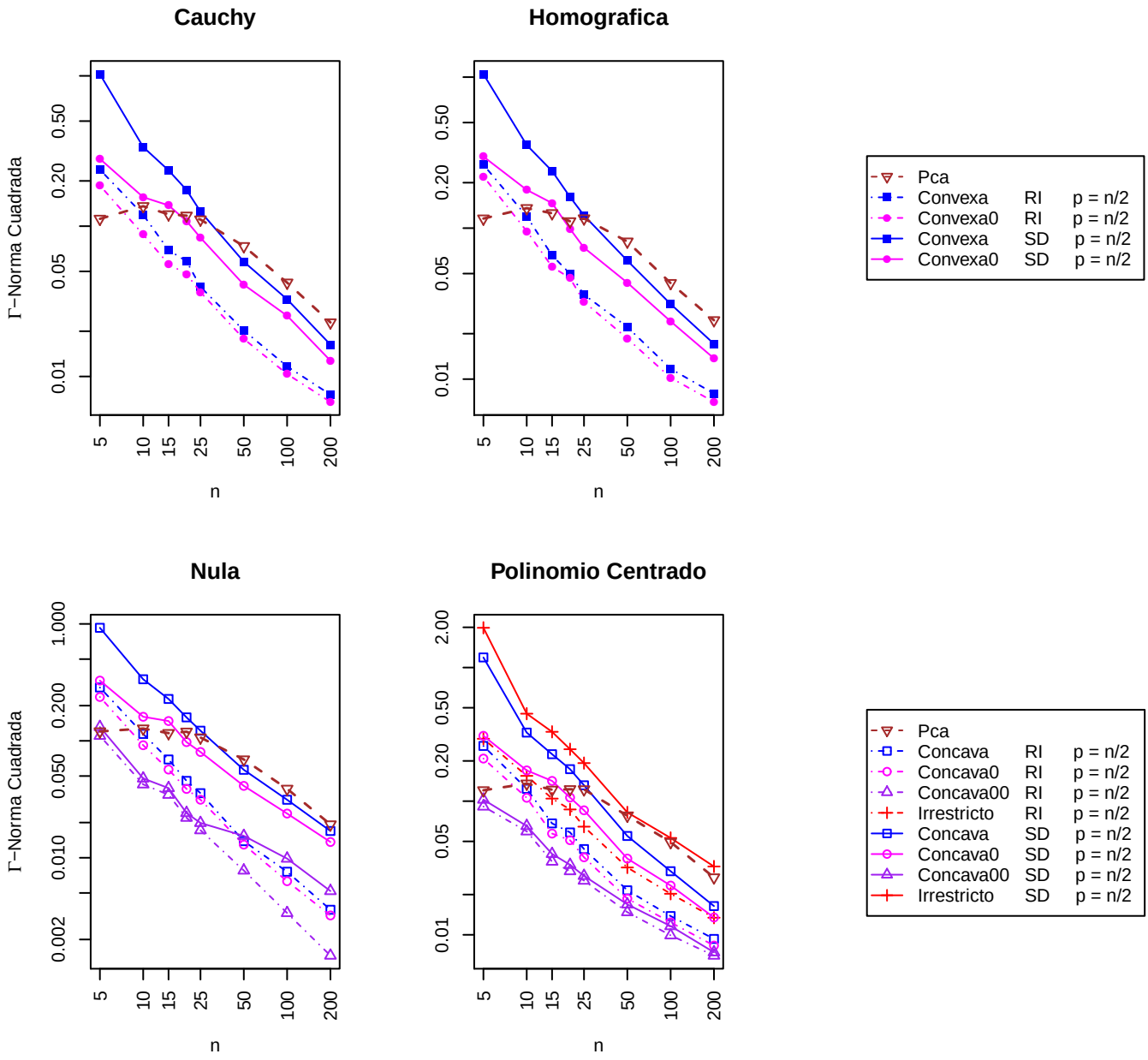

Figura 33: Curvas log-log de $\mathtt{E}\left(\|\widehat{b}_n - b\|^2_\Gamma\right)$ para estimadores con restricciones de concavidad y covariables Punch.

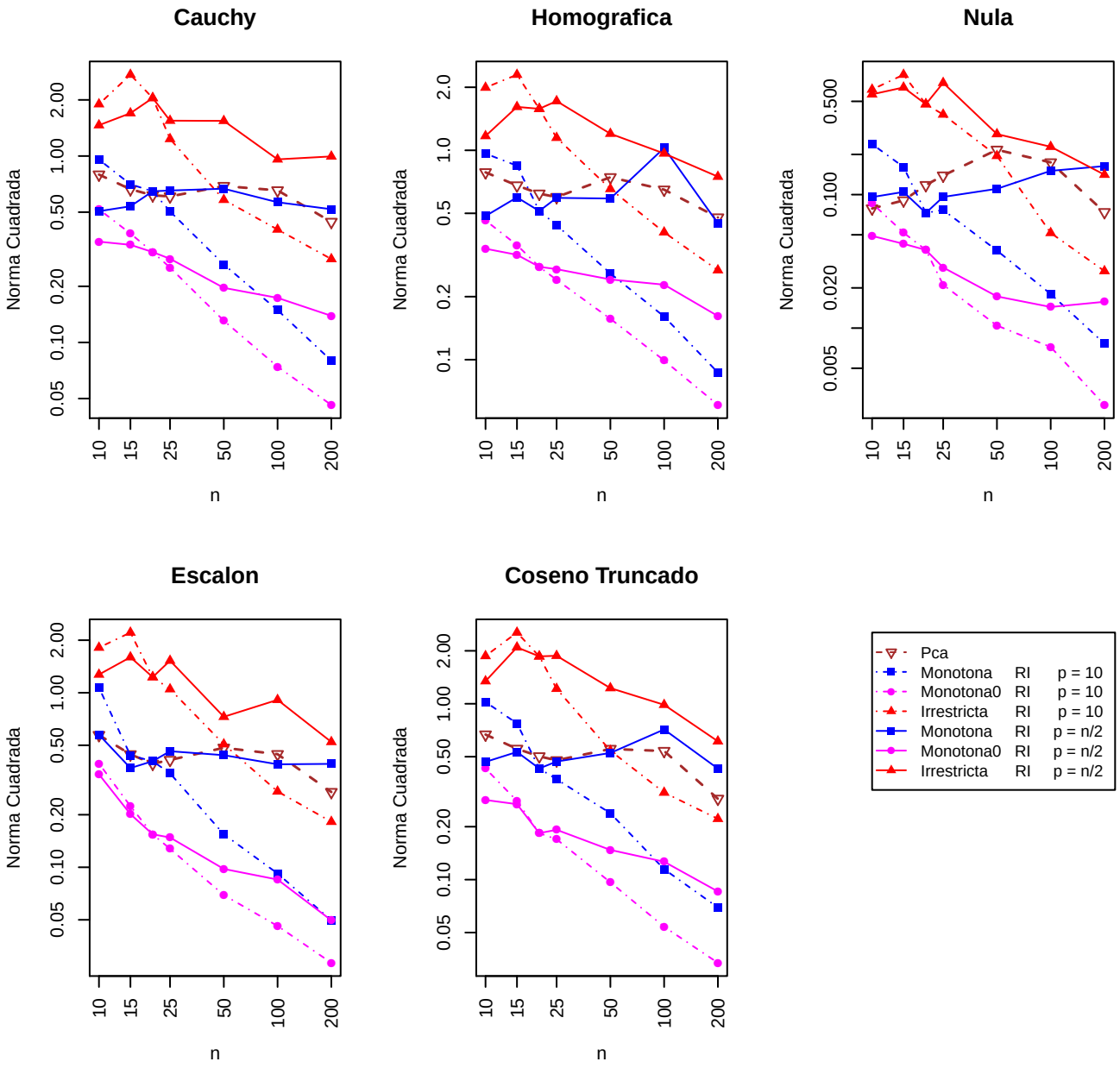

Figura 34: Curvas log-log de $\mathtt{E}\left(\|\widehat{b}_n - b\|^2\right)$ para estimadores con restricciones de monotonía y covariables Fourier $q = 1$.

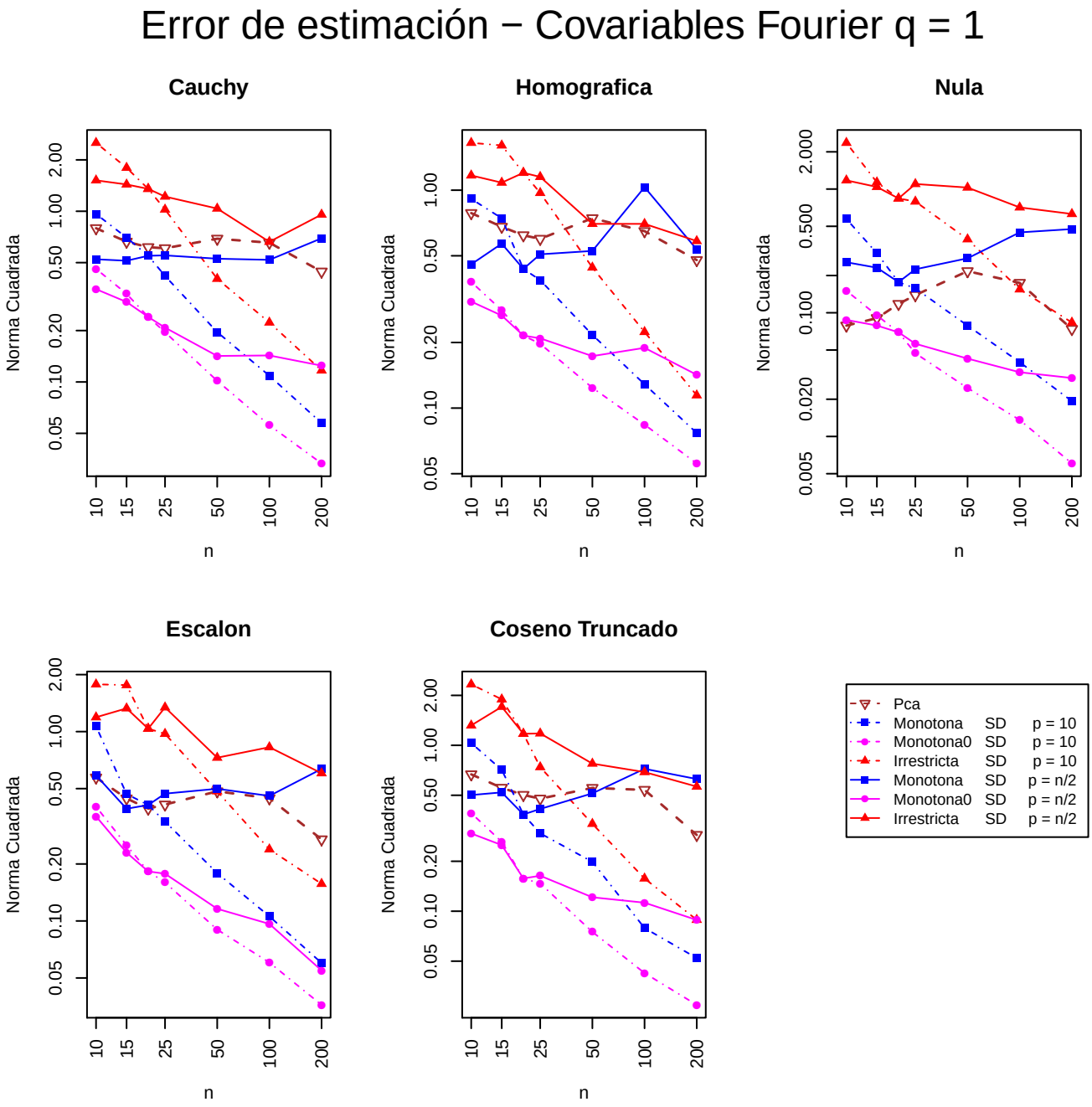

Figura 35: Curvas log-log de $\mathtt{E}\left(\|\widehat{b}_n - b\|^2\right)$ para estimadores con restricciones de monotonía y covariables Fourier $q = 1$.

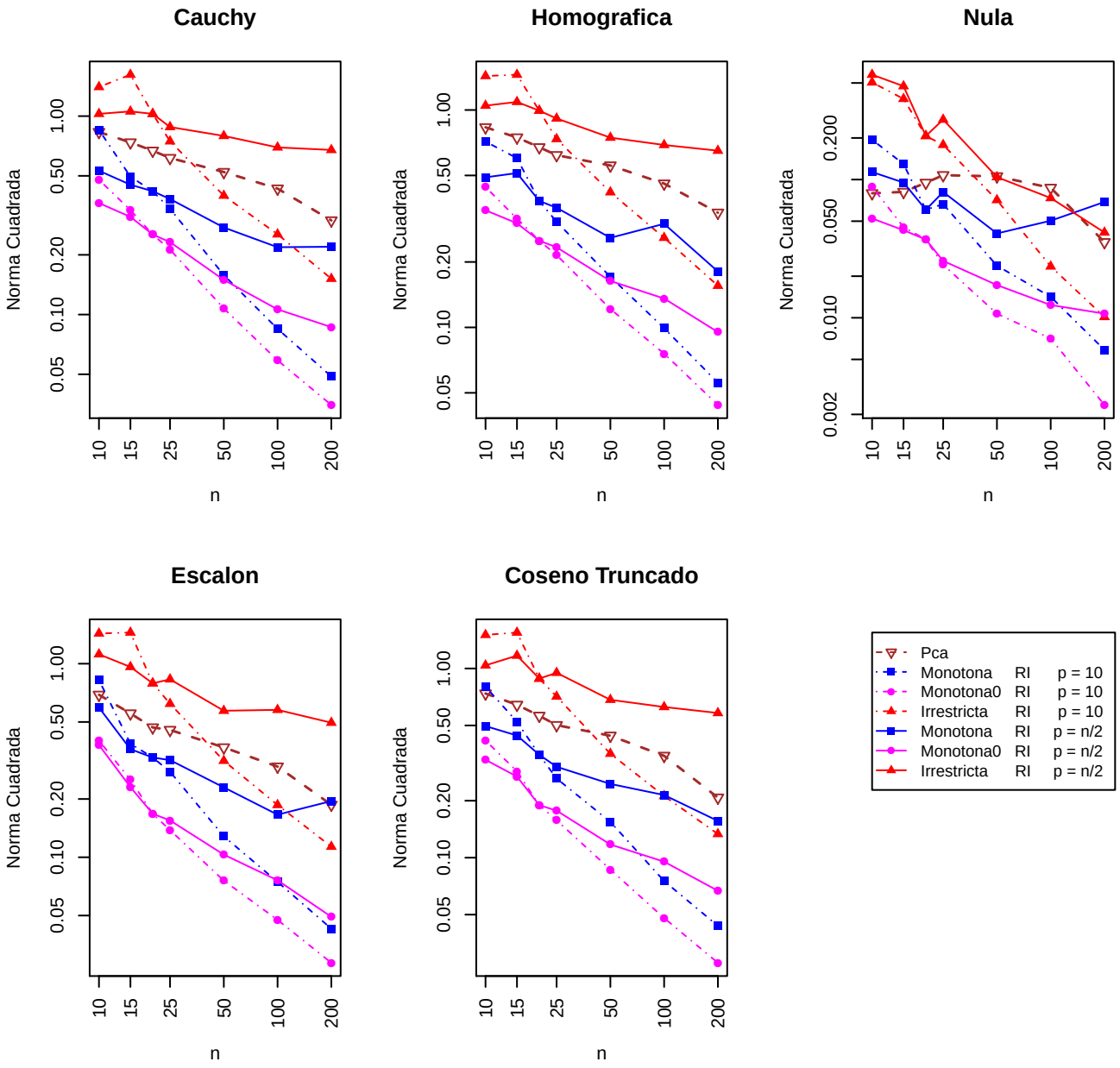

Figura 36: Curvas log-log de $\mathtt{E}\left(\|\widehat{b}_n - b\|^2\right)$ para estimadores con restricciones de monotonía y covariables Fourier $q = 0{,}2$.

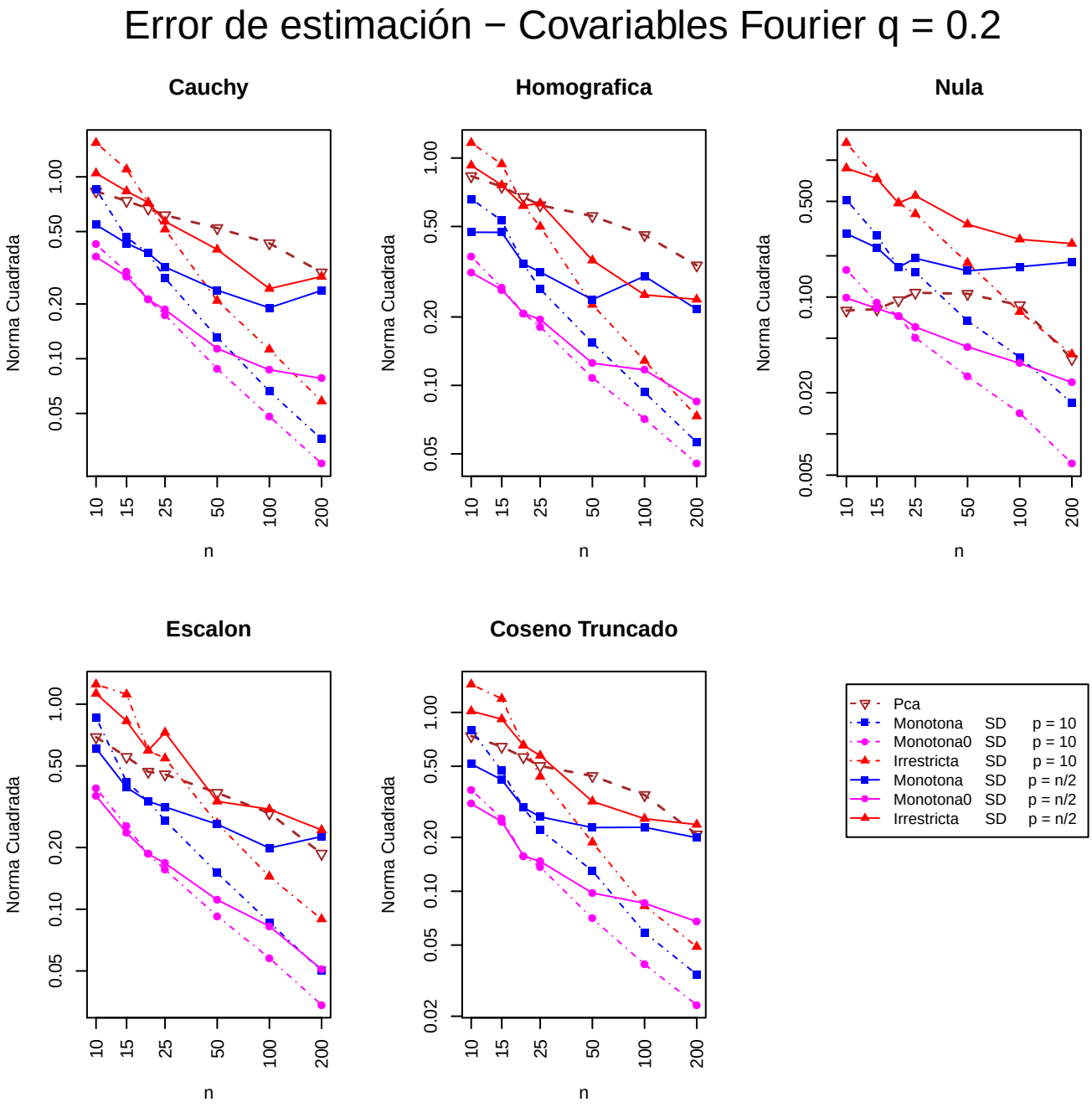

Figura 37: Curvas log-log de $\mathtt{E}\left(\|\widehat{b}_n - b\|^2\right)$ para estimadores con restricciones de monotonía y covariables Fourier $q = 0{,}2$.

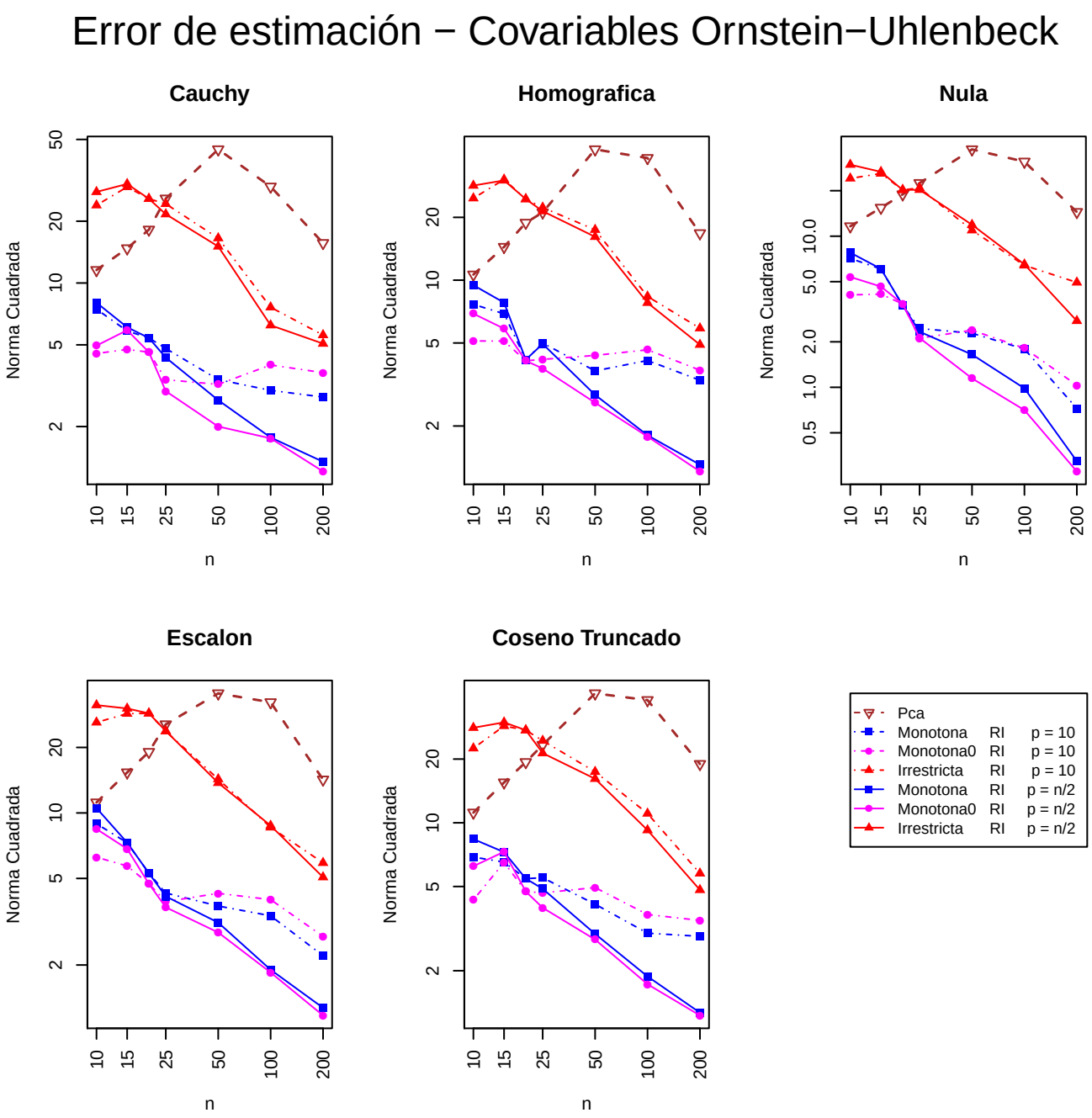

Figura 38: Curvas log-log de $\mathtt{E}\left(\|\widehat{b}_n - b\|^2\right)$ para estimadores con restricciones de monotonía y covariables Ornstein-Uhlenbeck.

n

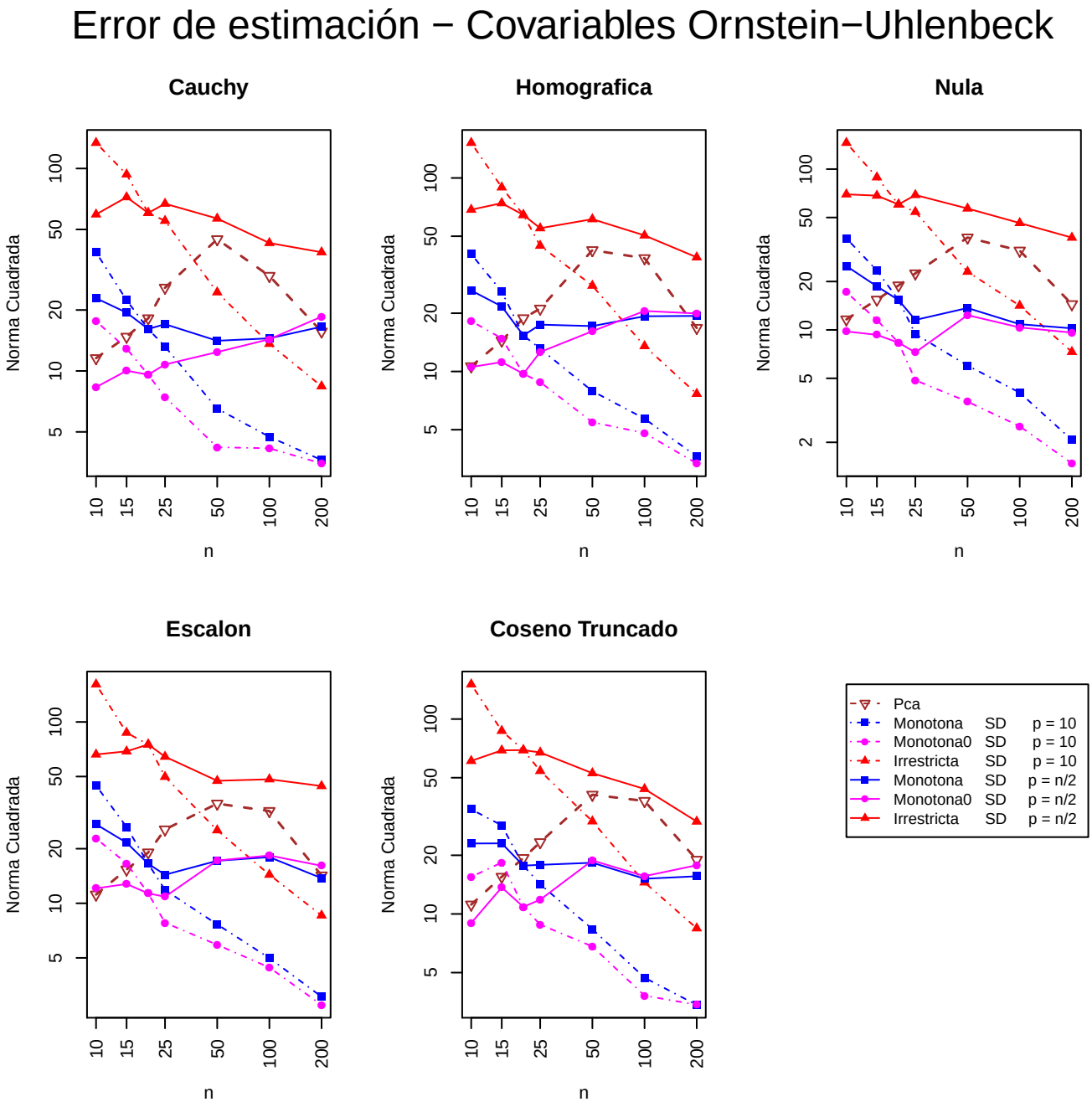

Figura 39: Curvas log-log de $\mathtt{E}\left(\|\widehat{b}_n - b\|^2\right)$ para estimadores con restricciones de monotonía y covariables Ornstein-Uhlenbeck.

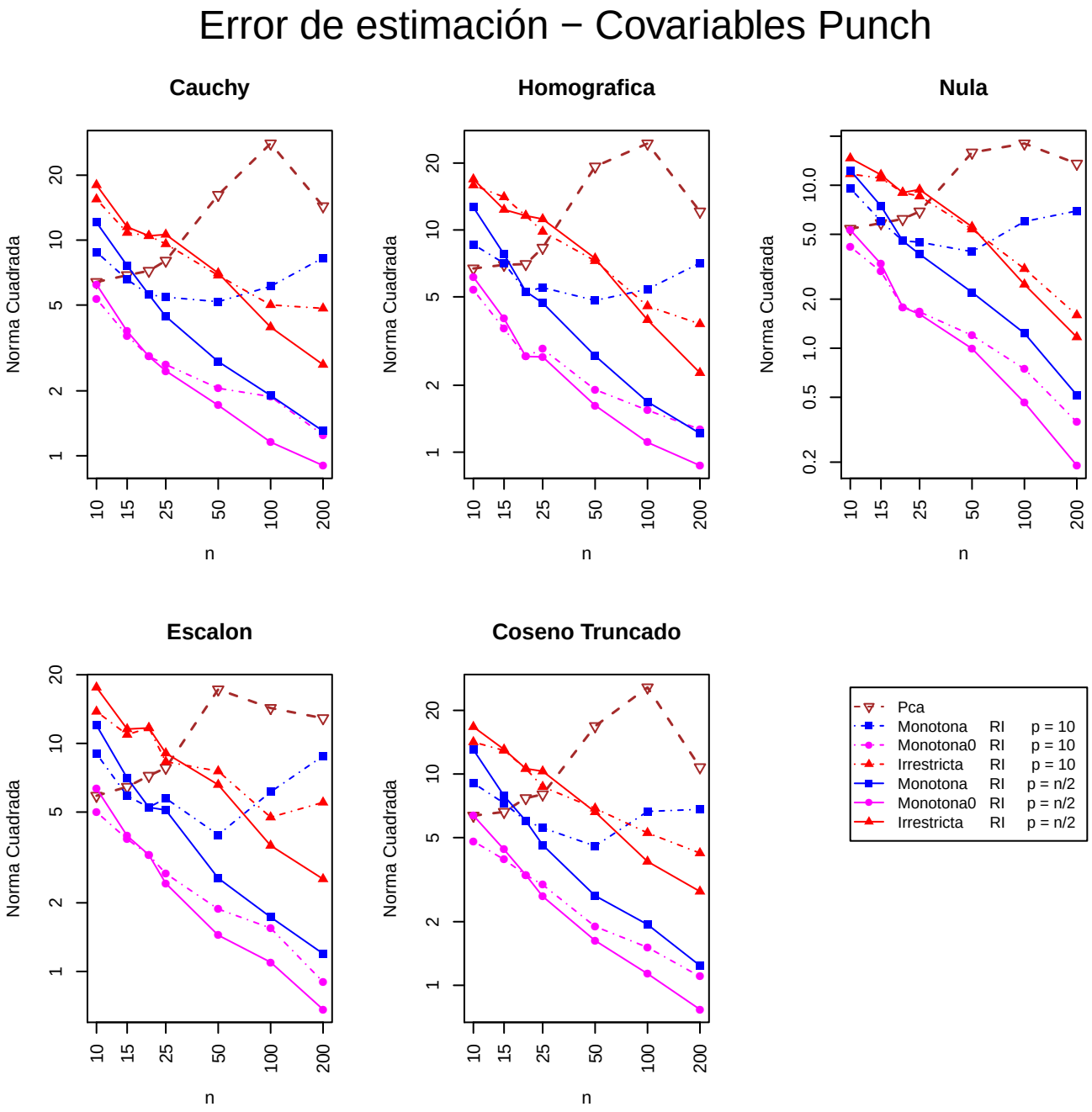

Figura 40: Curvas log-log de $\mathtt{E}\left(\|\widehat{b}_n - b\|^2\right)$ para estimadores con restricciones de monotonía y covariables Punch.

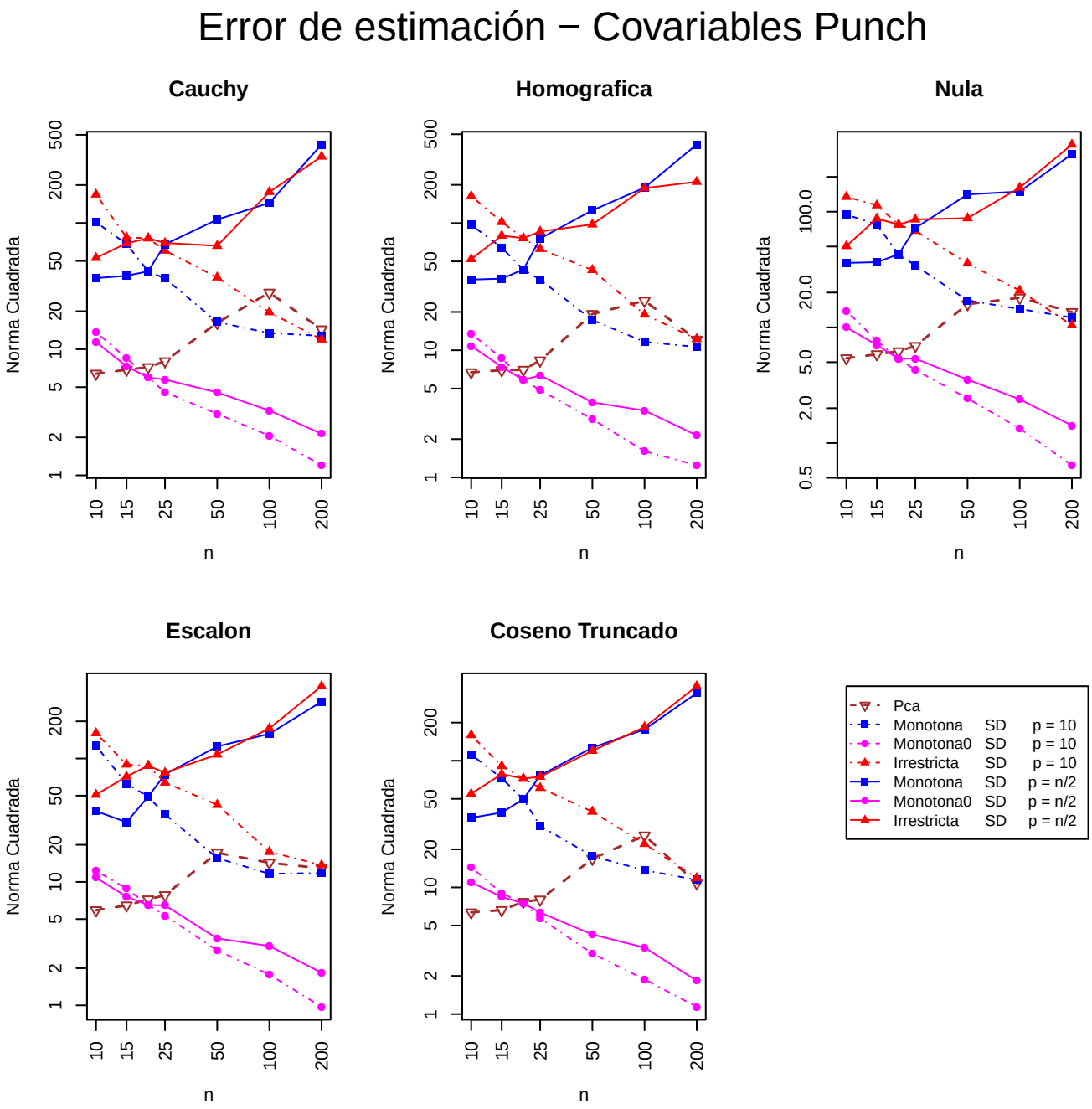

Figura 41: Curvas log-log de $\mathtt{E}\left(\|\widehat{b}_n - b\|^2\right)$ para estimadores con restricciones de monotonía y covariables Punch.

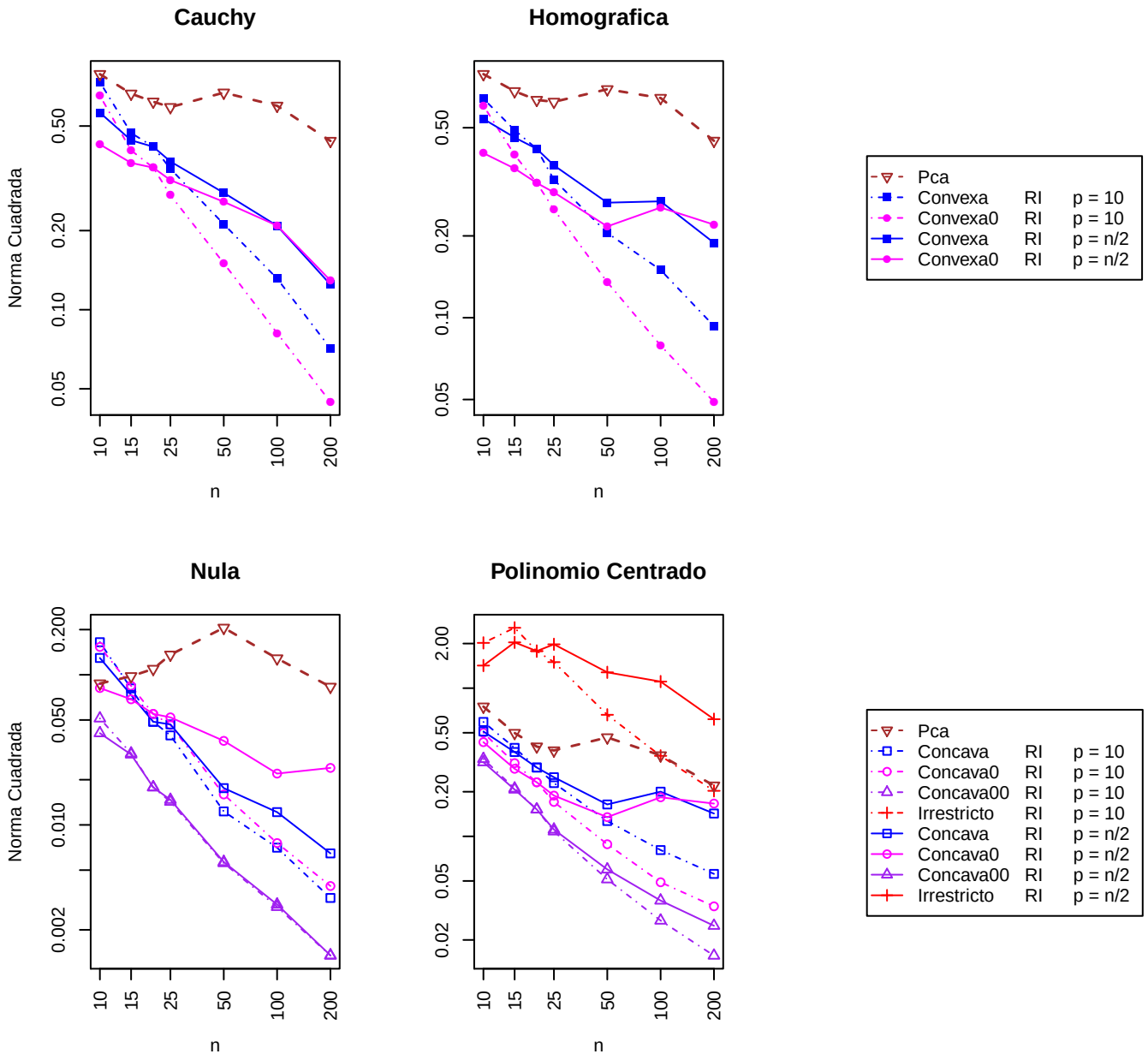

Figura 42: Curvas log-log de $\mathtt{E}\left(\|\widehat{b}_n - b\|^2\right)$ para estimadores con restricciones de concavidad y covariables Fourier $q = 1$.

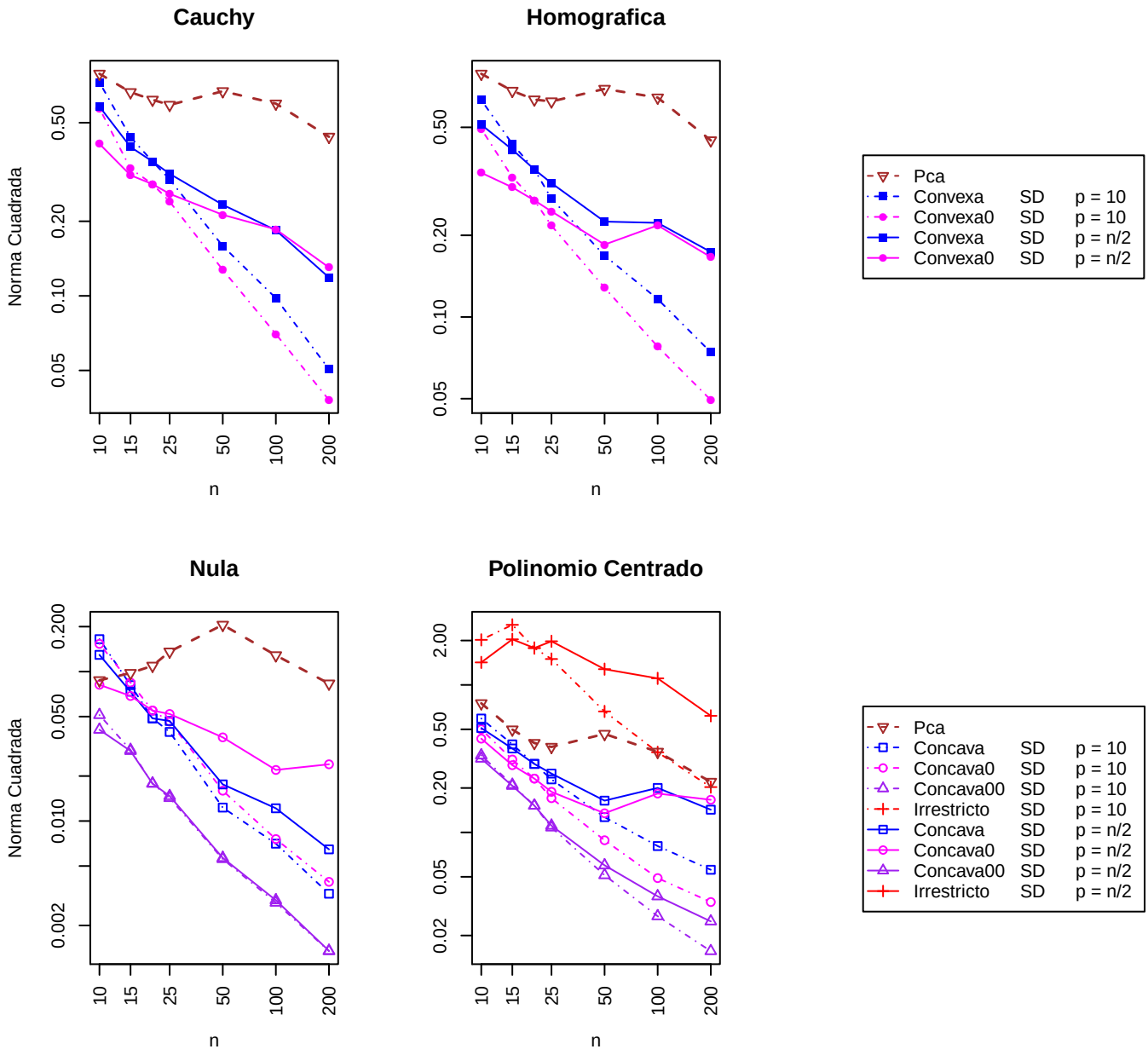

Figura 43: Curvas log-log de $\mathtt{E}\left(\|\widehat{b}_n - b\|^2\right)$ para estimadores con restricciones de concavidad y covariables Fourier $q = 1$.

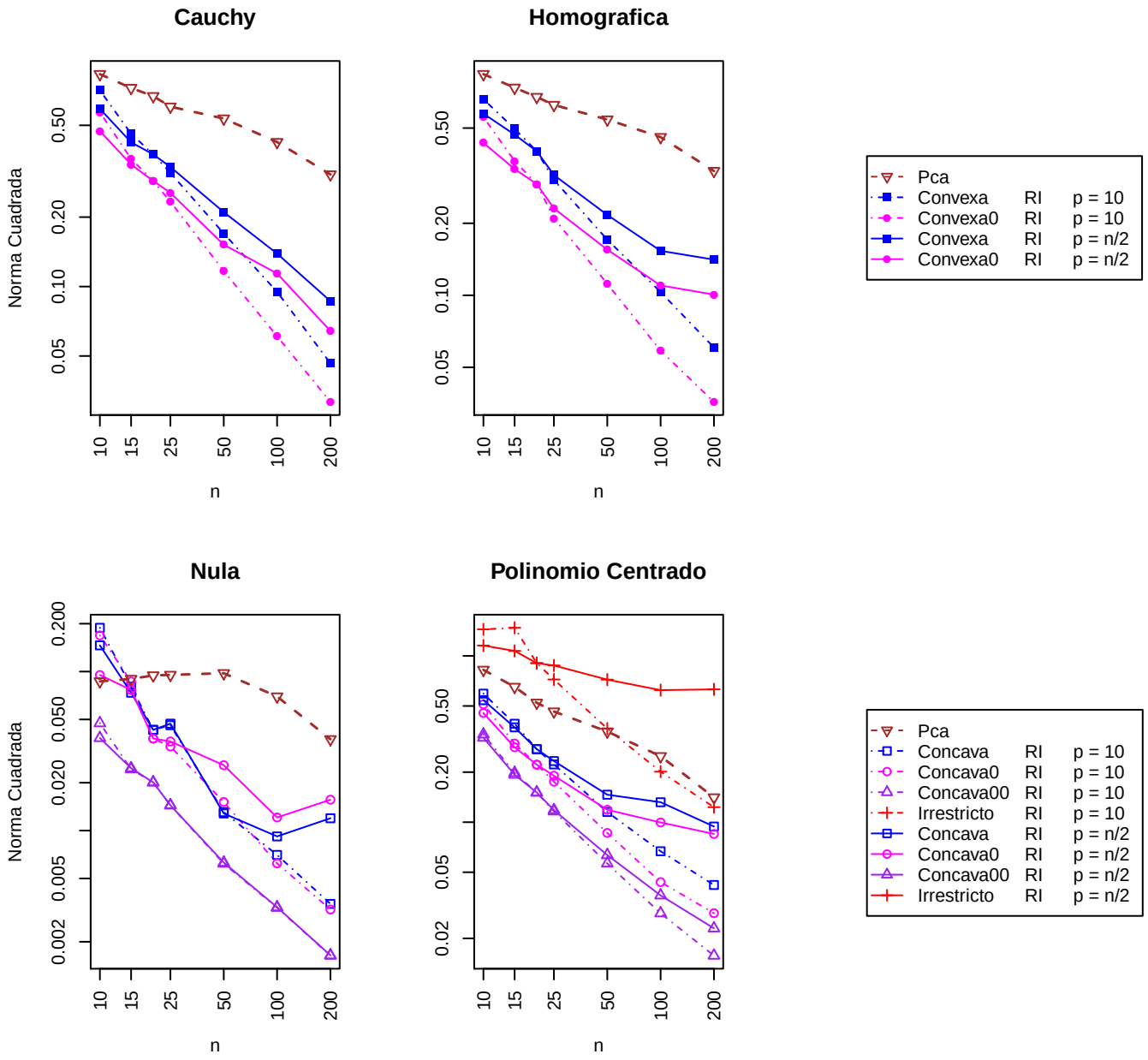

Figura 44: Curvas log-log de $\mathtt{E}\left(\|\widehat{b}_n - b\|^2\right)$ para estimadores con restricciones de concavidad y covariables Fourier $q = 0{,}2$.

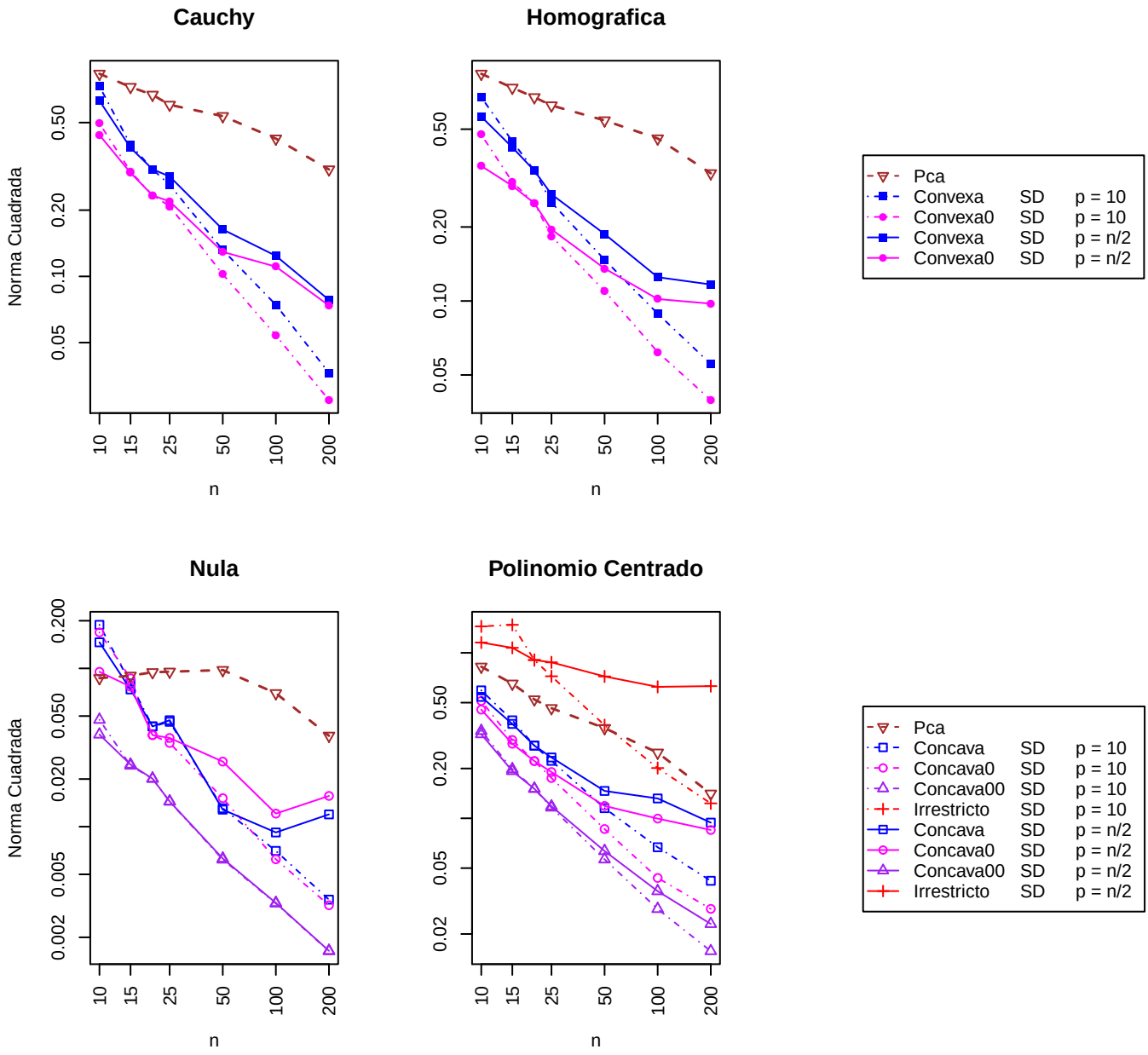

Figura 45: Curvas log-log de $\mathtt{E}\left(\|\widehat{b}_n - b\|^2\right)$ para estimadores con restricciones de concavidad y covariables Fourier $q = 0{,}2$.

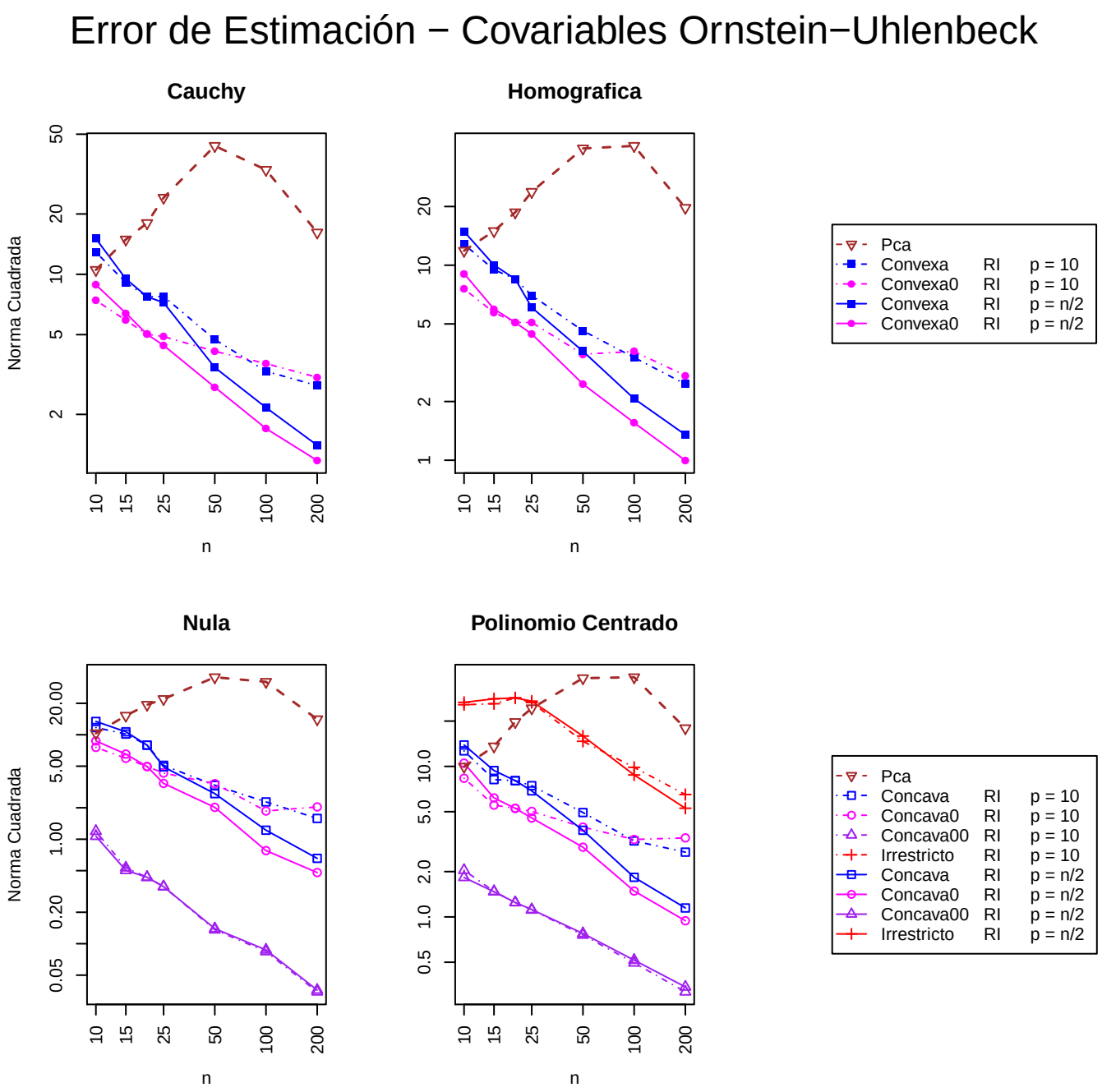

Figura 46: Curvas log-log de $\mathtt{E}\left(\|\widehat{b}_n - b\|^2\right)$ para estimadores con restricciones de concavidad y covariables Ornstein-Uhlebenck.

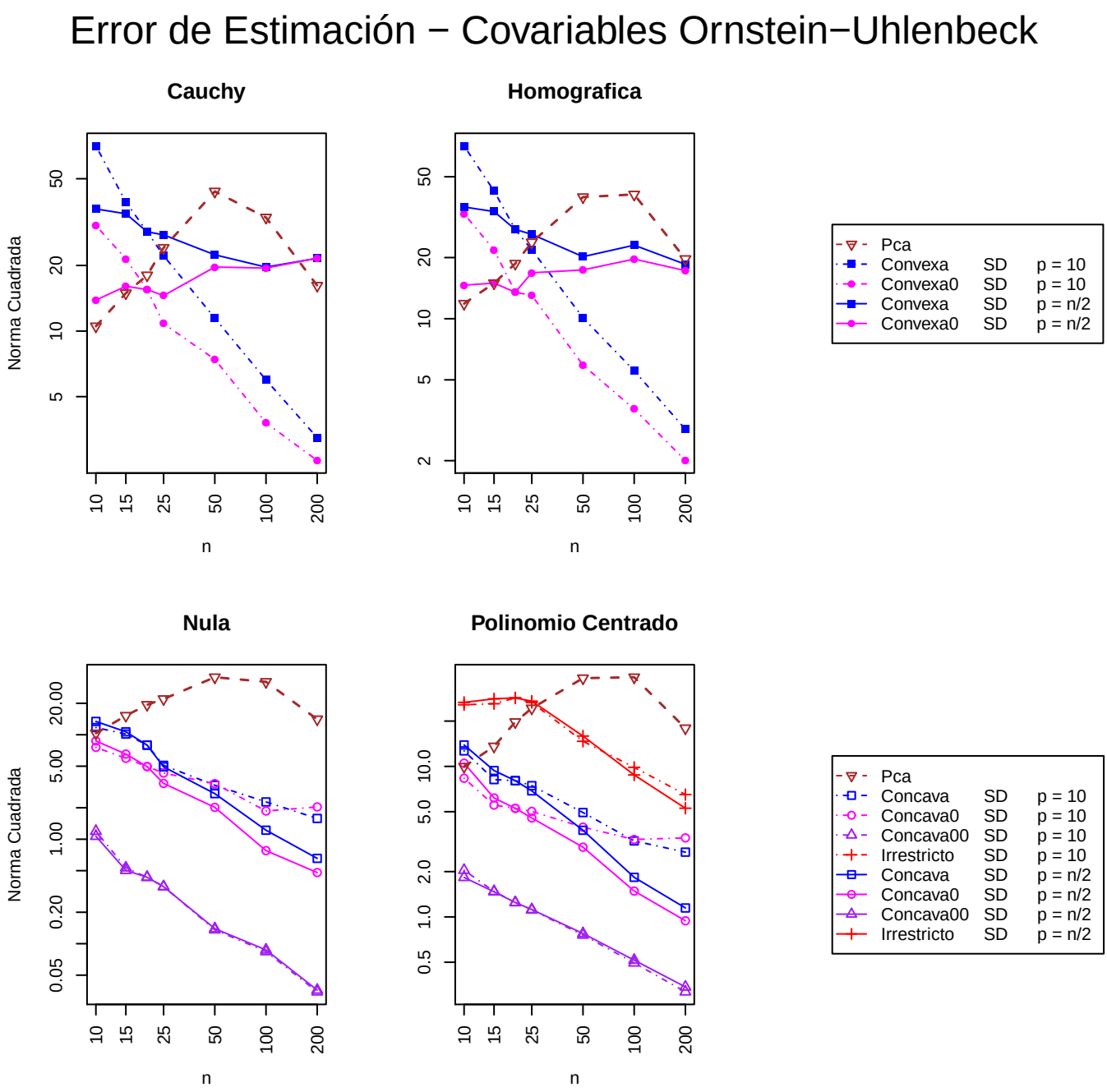

Figura 47: Curvas log-log de $\mathtt{E}\left(\|\widehat{b}_n - b\|^2\right)$ para estimadores con restricciones de concavidad y covariables Ornstein-Uhlebenck.

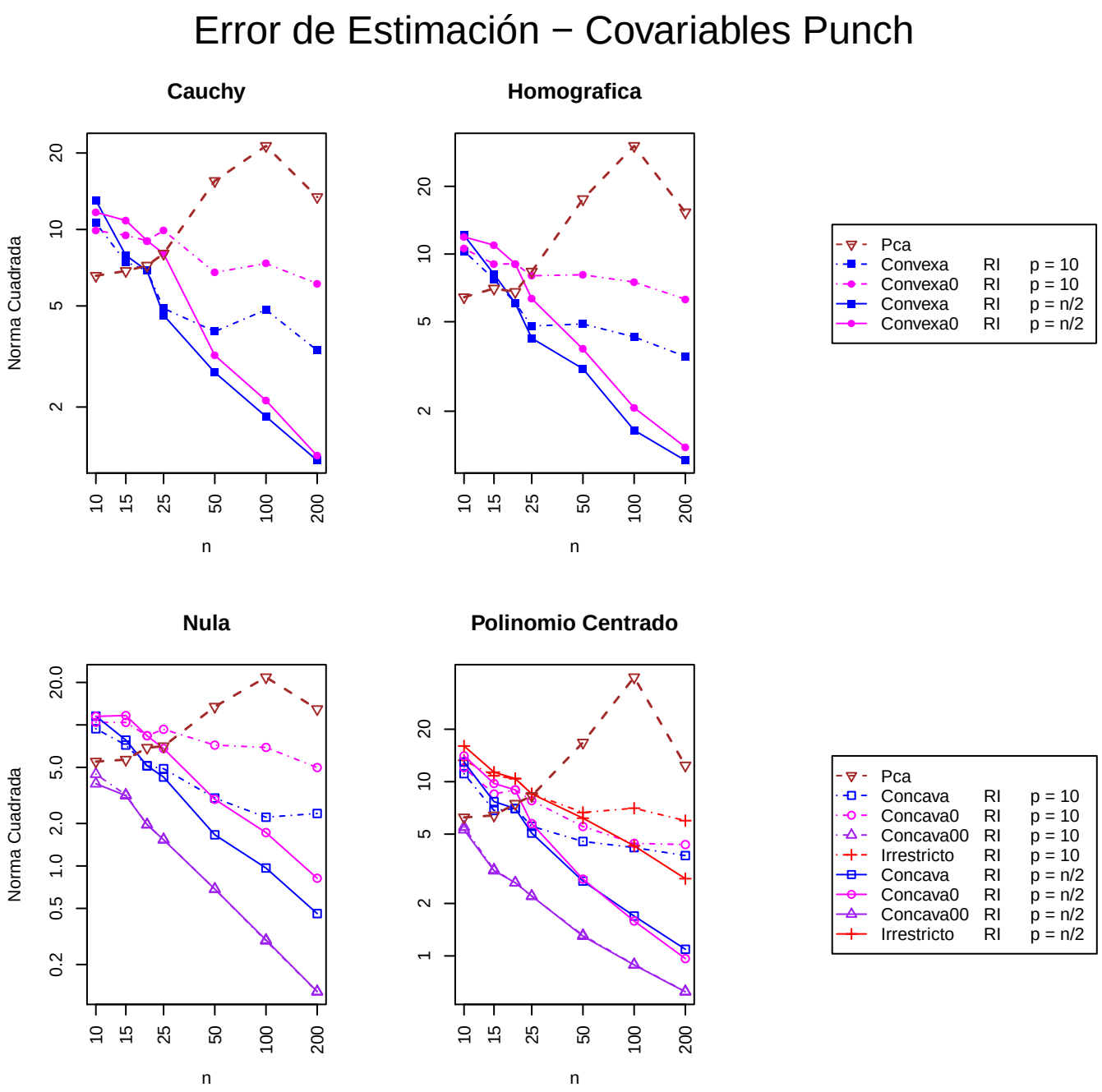

Figura 48: Curvas log-log de $\mathtt{E}\left(\|\widehat{b}_n - b\|^2\right)$ para estimadores con restricciones de concavidad y covariables Punch.

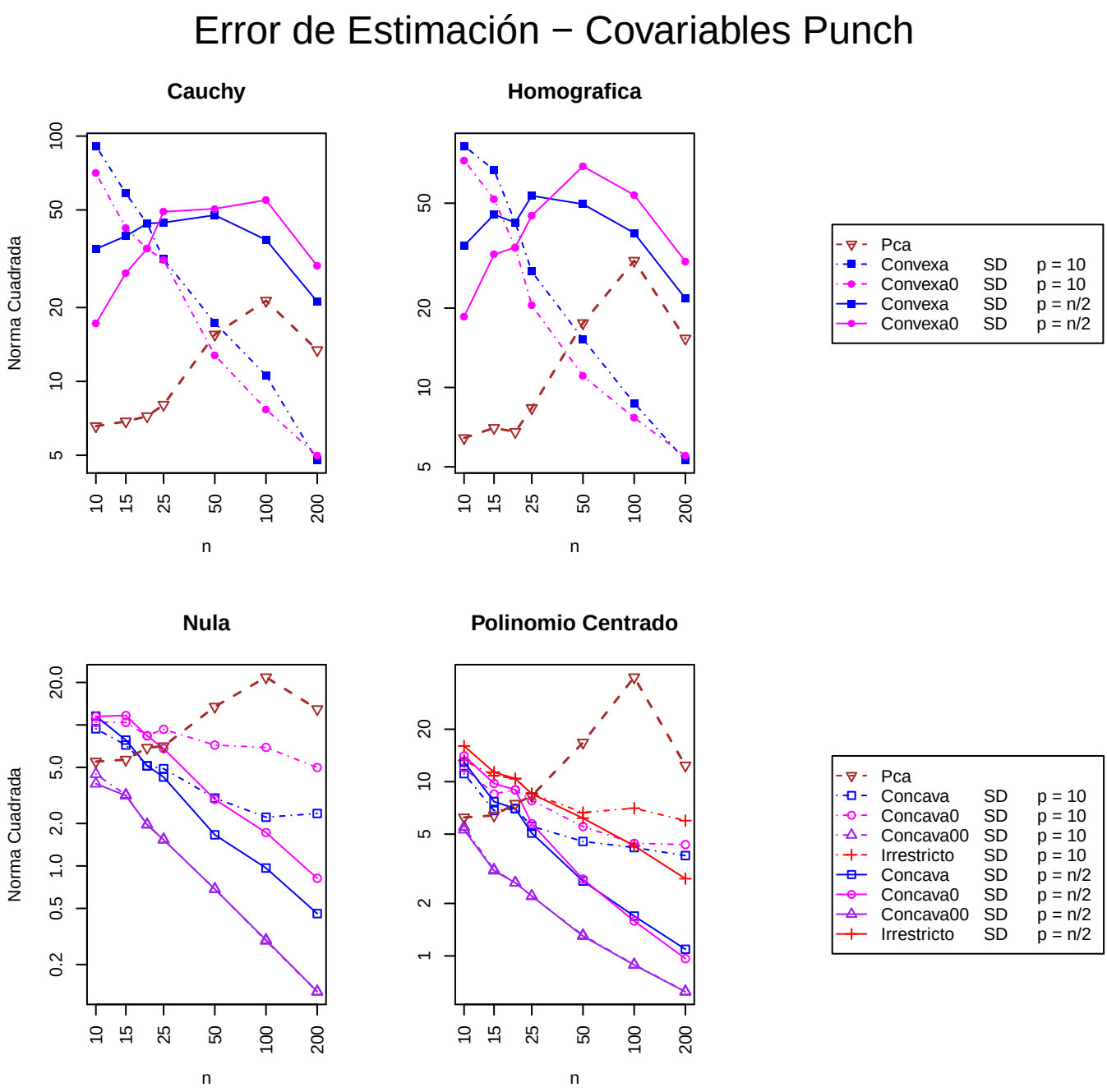

Figura 49: Curvas log-log de $\mathtt{E}\left(\|\widehat{b}_n - b\|^2\right)$ para estimadores con restricciones de concavidad y covariables Punch.

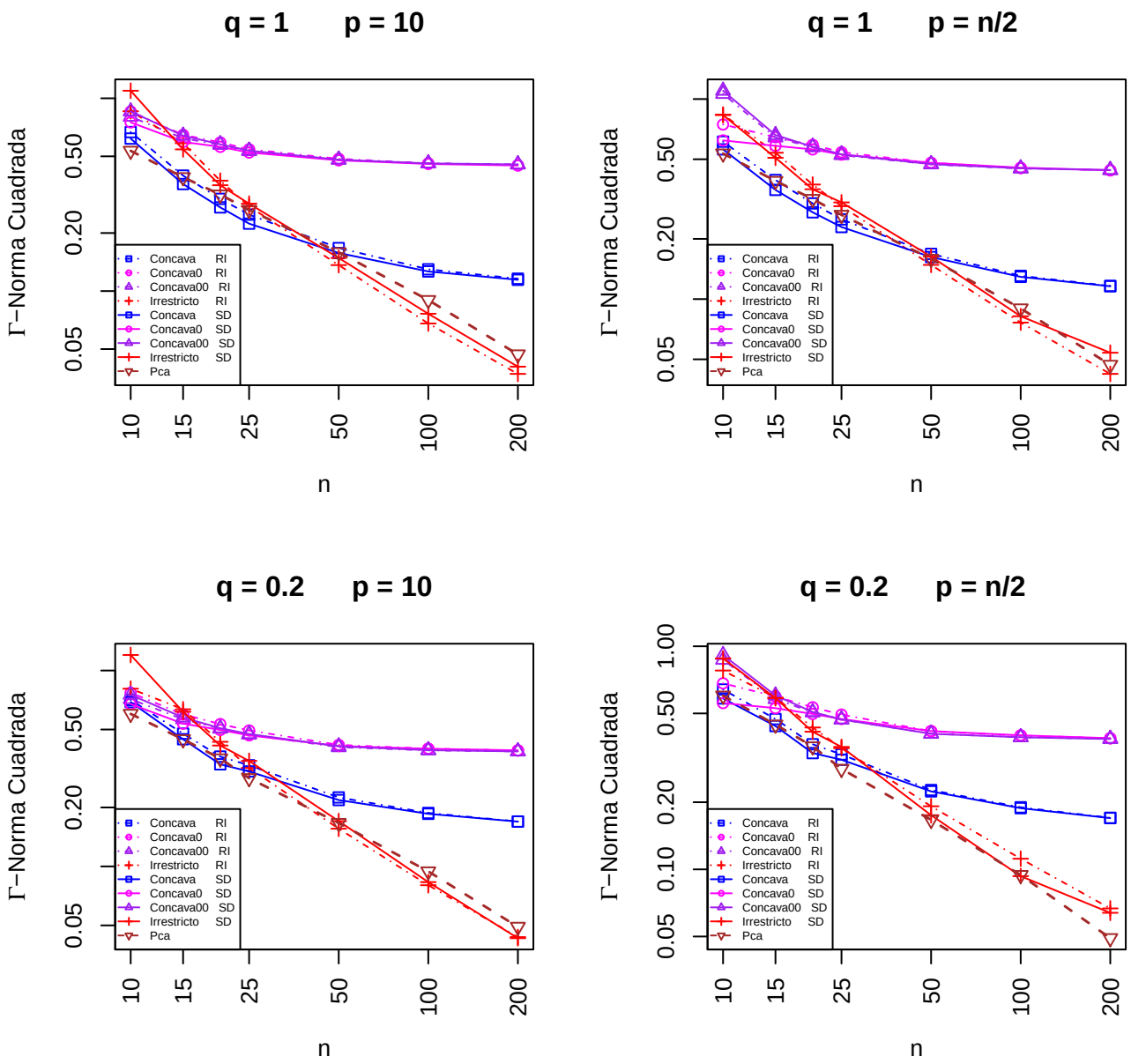

Figura 50: Curvas log-log de $\mathtt{E}\left(\|\widehat{b}_n - b\|^2_\Gamma\right)$ para la función Pico con covariables Fourier $q = 1$ y $q = 0{,}2$. La función Pico no cumple con la restricciones de forma bajo la que se la estima.

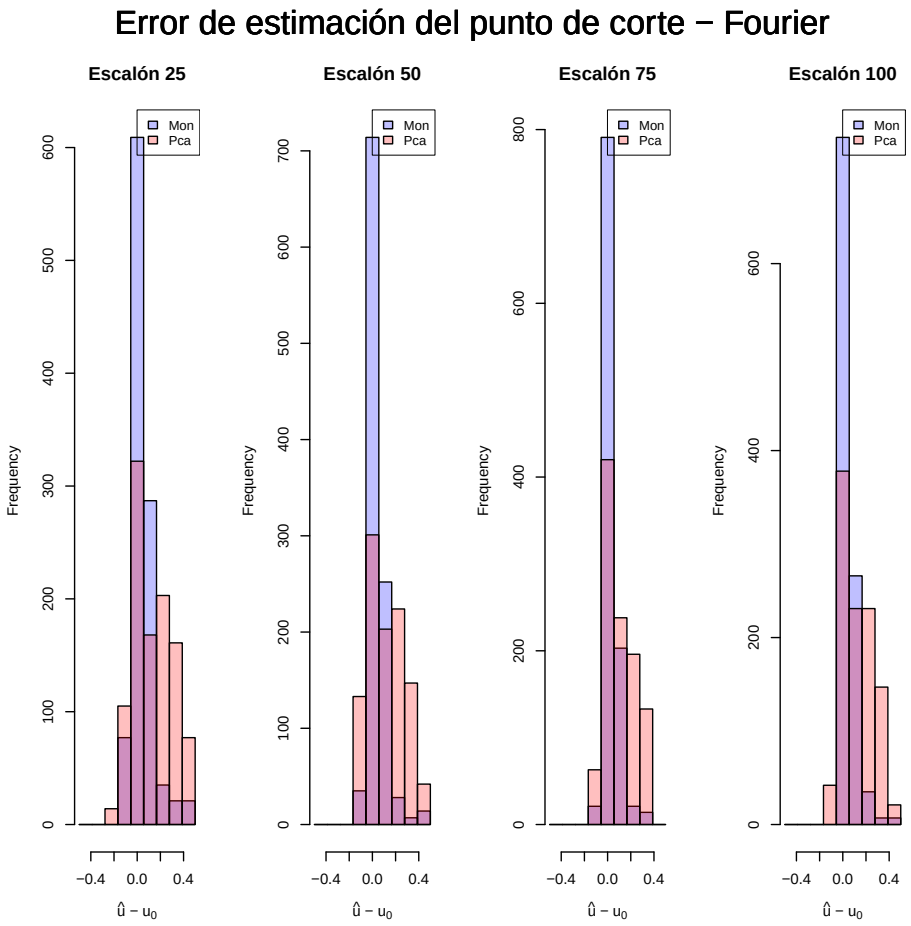

Figura 51: Errores de estimación del punto de corte para la función Escalón con covariables Fourier $q = 1$.

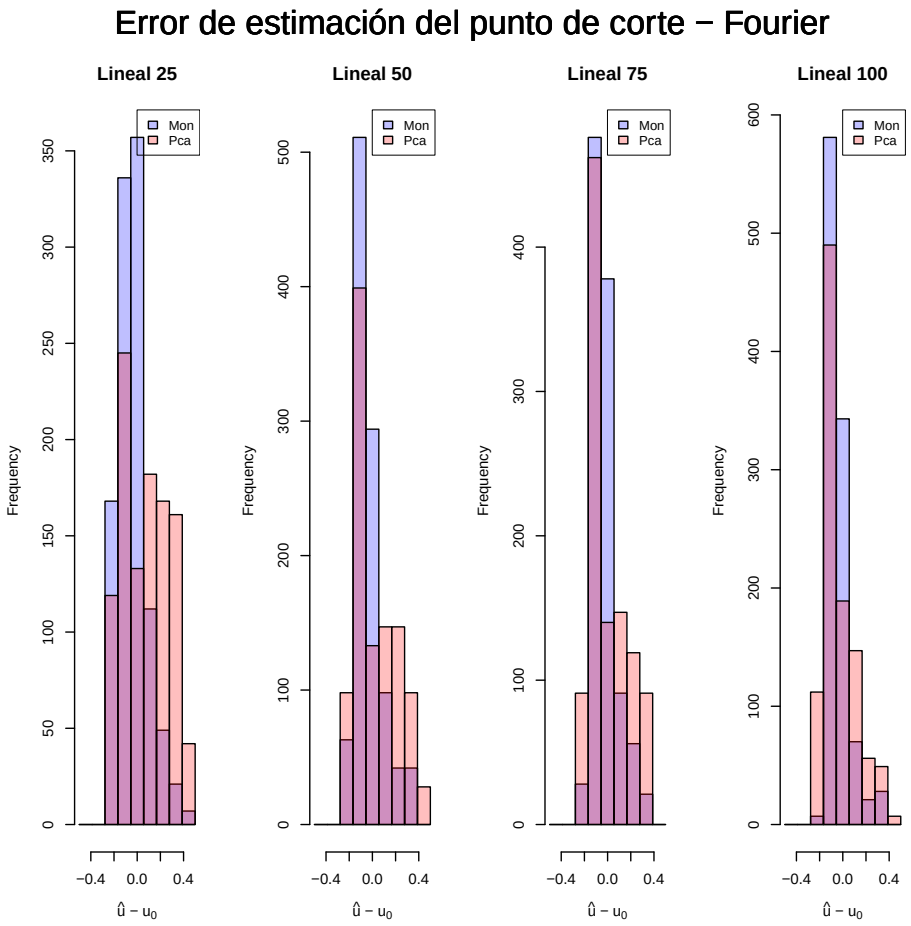

Figura 52: Errores de estimación del punto de corte para la función Lineal con covariables Fourier $q = 1$.

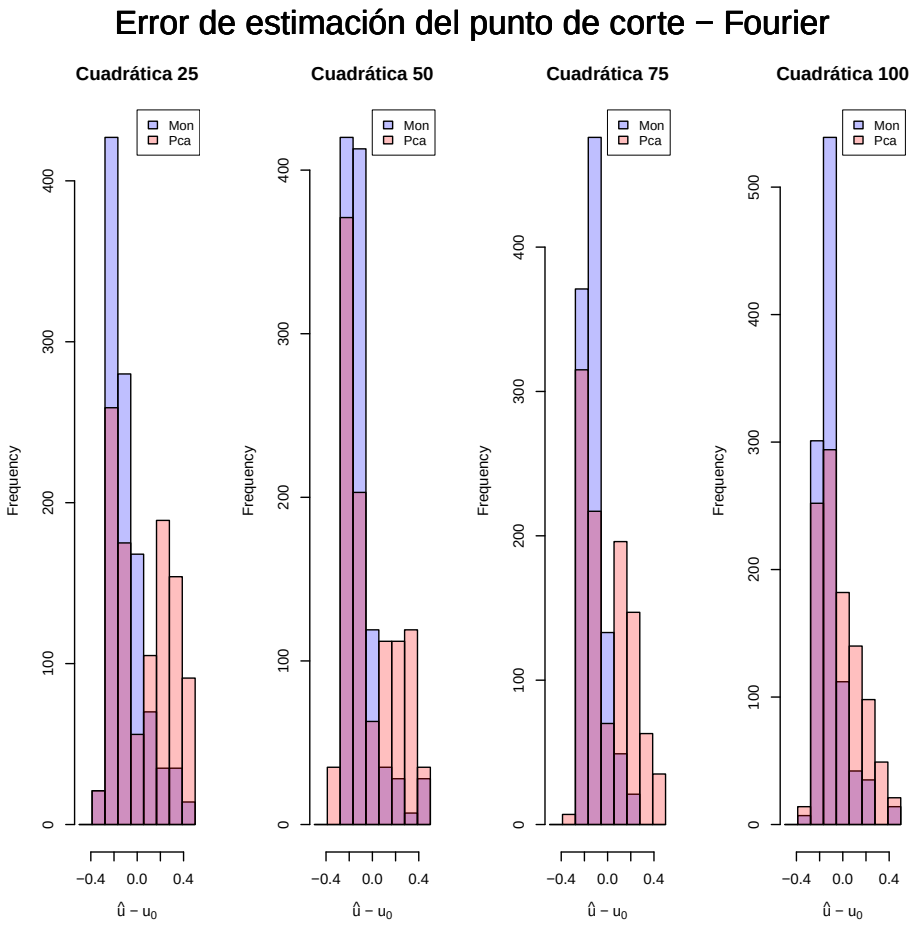

Figura 53: Errores de estimación del punto de corte para la función Cuadrática con covariables Fourier $q = 1$.

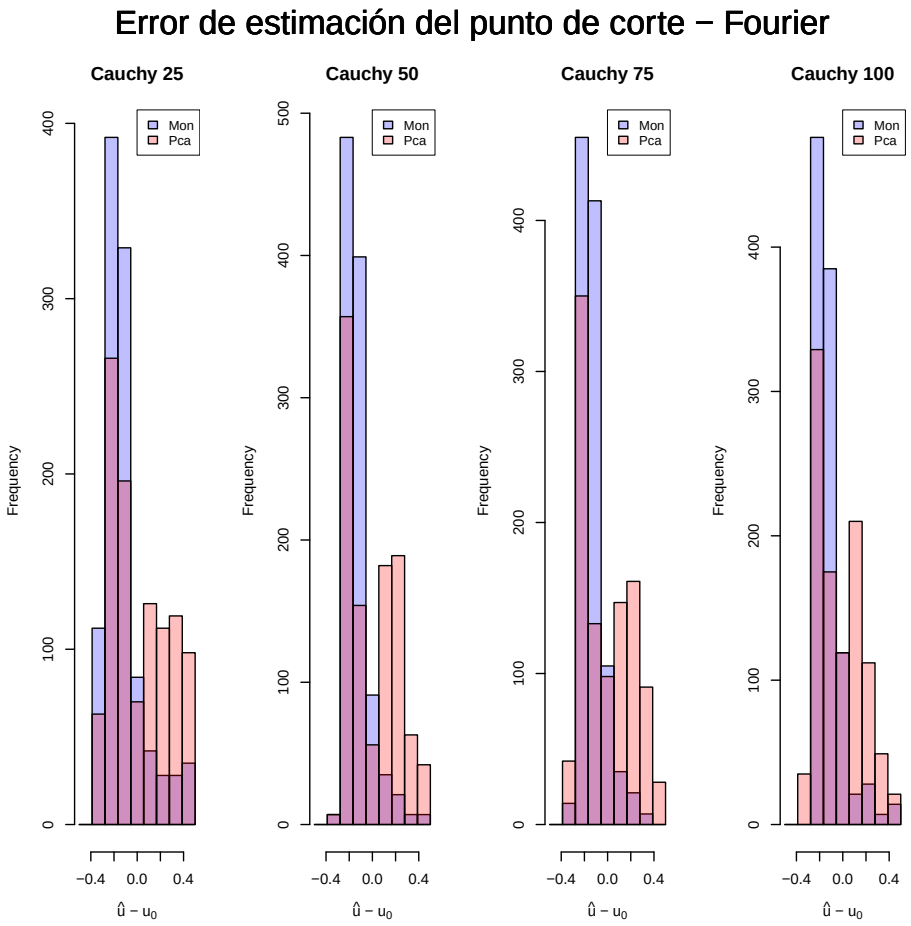

Figura 54: Errores de estimación del punto de corte para la Función Cauchy con covariables Fourier $q = 1$.

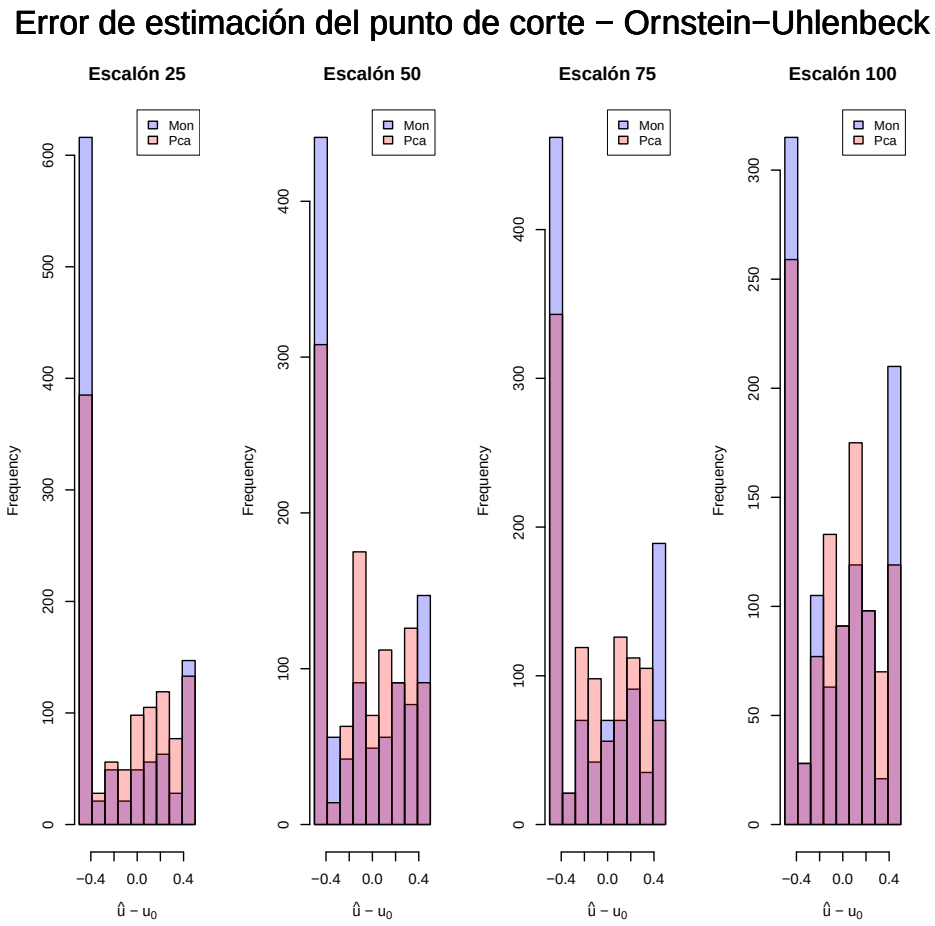

Figura 55: Errores de estimación del punto de corte para la función Escalón con covariables Ornstein-Uhlenbeck.

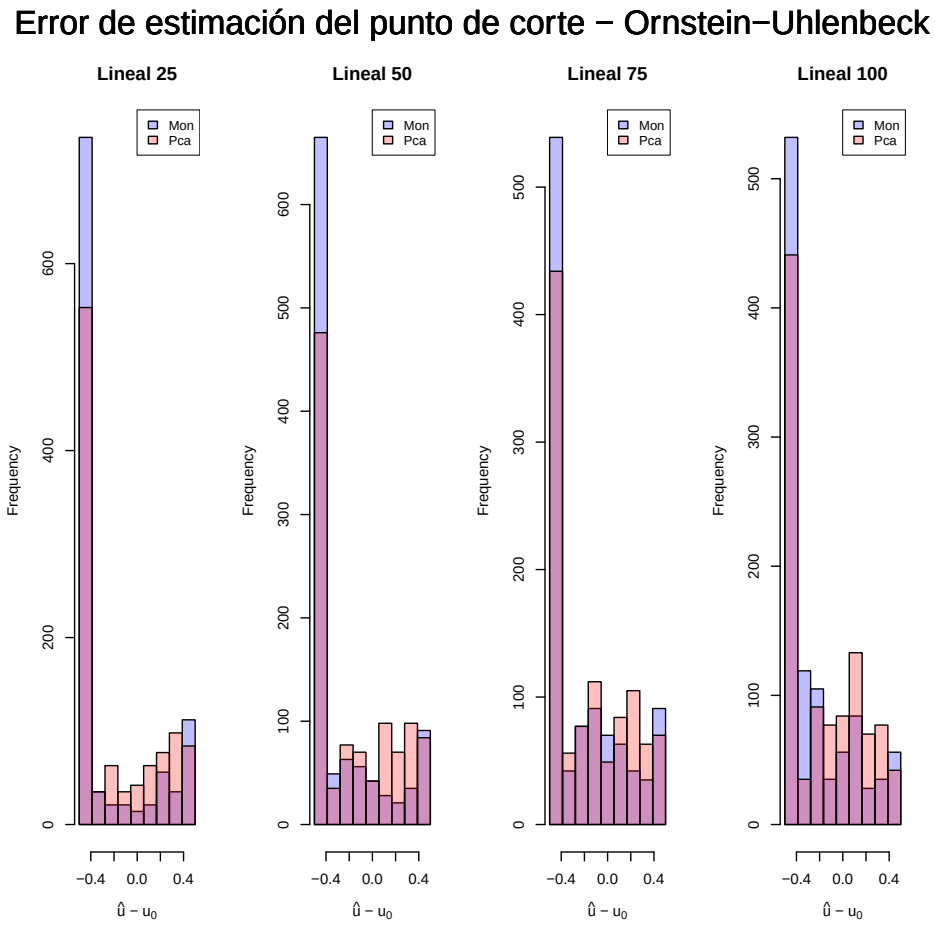

Figura 56: Errores de estimación del punto de corte para la función Lineal con covariables Ornstein-Uhlenbeck.

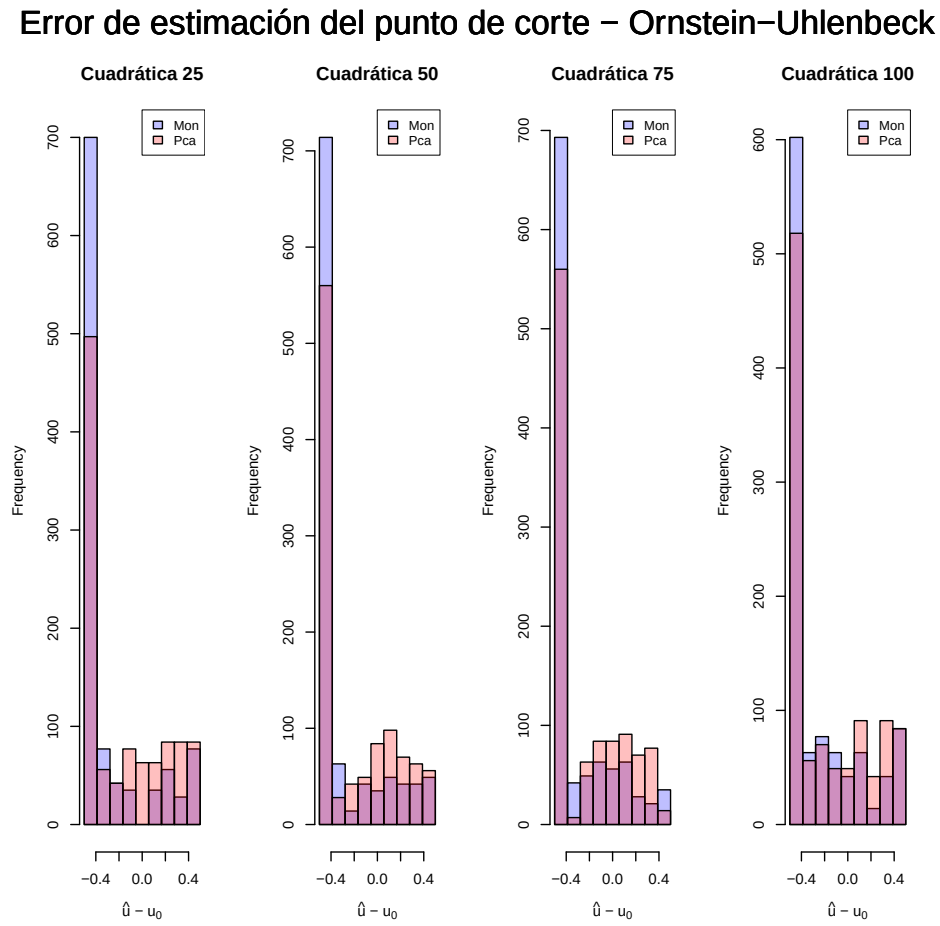

Figura 57: Errores de estimación del punto de corte para la función Cuadrática con covariables Ornstein-Uhlenbeck.

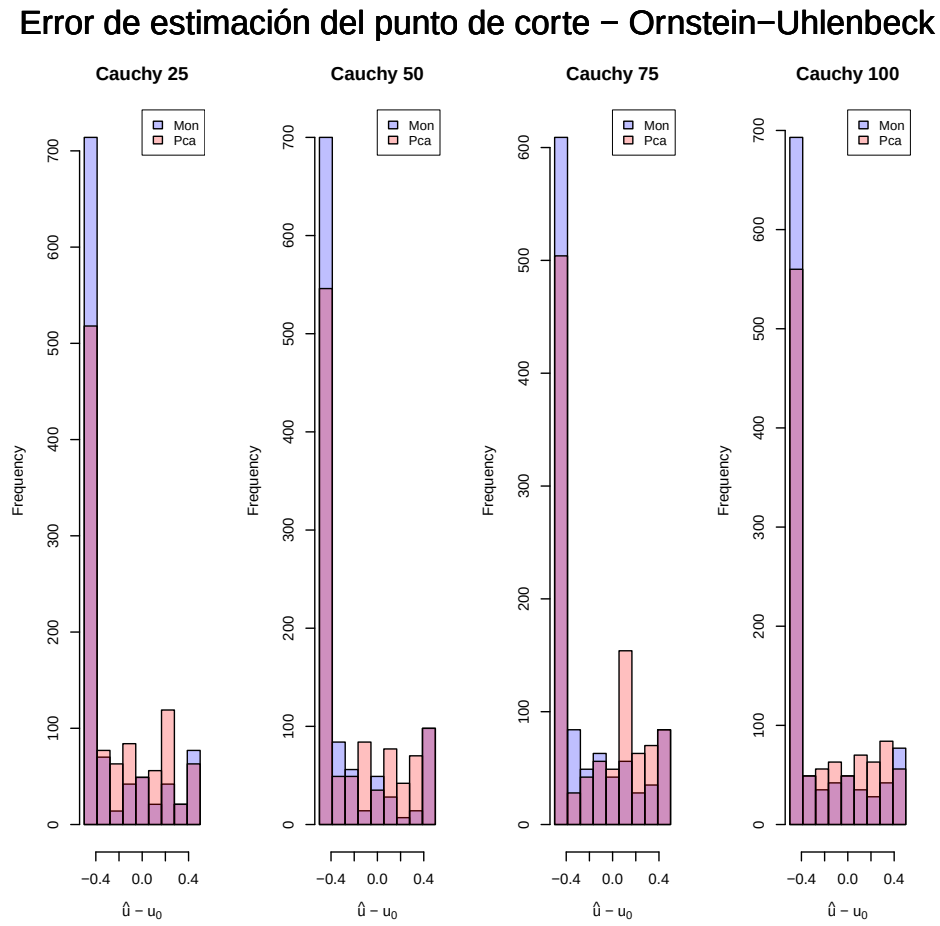

Figura 58: Errores de estimación del punto de corte para la función Cauchy con covariables Ornstein-Uhlenbeck.

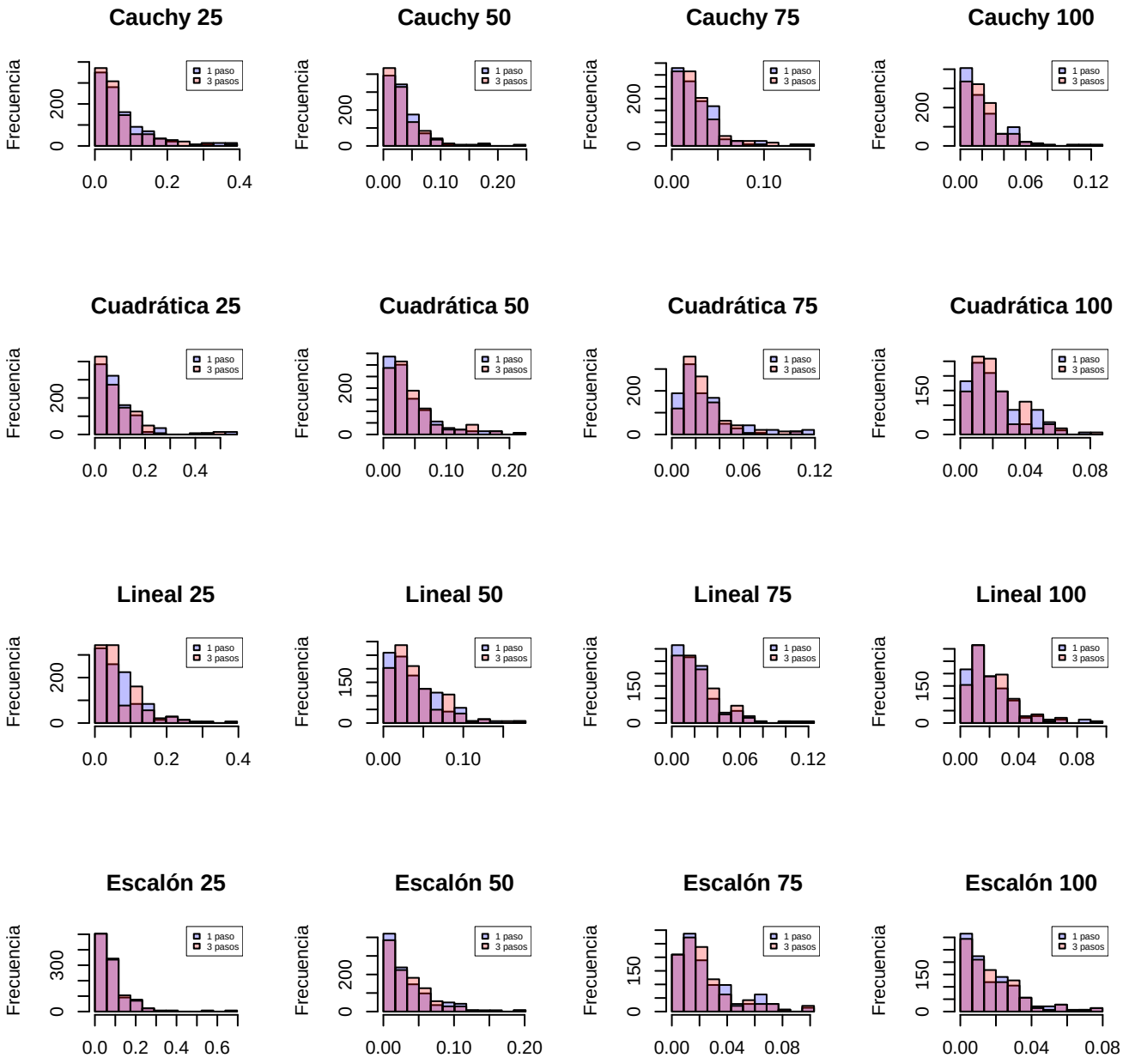

Figura 59: Errores de estimación medidos en Γ-Norma para los estimadores a un paso y tres pasos con inicial Monótono0.

CAUCHY										
Metodología			Cantidad de covariables							
Restricción	Penalidad	p	5	10	15	20	25	50	100	200
Monótona	RI	10	0.29	0.12	0.12	0.10	0.17	0.11	0.05	0.07
Monótona	RI	n/2	0.04	**0**	0.10	0.10	0.18	0.18	0.15	0.20
Monótona	SD	10	0.30	0.04	**0**	**0**	0.03	0.07	**0**	**0**
Monótona	SD	n/2	**0**	0.01	0.01	**0**	**0**	**0**	0.06	0.19
Convexa	RI	10	0.60	0.42	0.37	0.41	0.38	0.36	0.25	0.16
Convexa	RI	n/2	0.37	0.27	0.36	0.41	0.39	0.39	0.30	0.21
Convexa	SD	10	1.53	0.22	0.13	0.19	0.25	0.27	0.20	0.11
Convexa	SD	n/2	0.48	0.20	0.16	0.19	0.21	0.22	0.24	0.21
Valor Mínimo ECM			0.72	0.34	0.21	0.15	0.12	0.06	0.03	0.02

HOMOGRAFICA										
Metodología			Cantidad de covariables							
Restricción	Penalidad	p	5	10	15	20	25	50	100	200
Monótona	RI	10	0.31	0.26	0.16	0.17	0.15	0.17	0.13	0.12
Monótona	RI	n/2	0.07	0.14	0.13	0.17	0.16	0.18	0.20	0.22
Monótona	SD	10	0.53	0.05	**0**	**0**	0.01	0.02	0.04	0.09
Monótona	SD	n/2	**0**	**0**	**0**	**0**	**0**	**0**	0.10	0.16
Convexa	RI	10	1.00	0.56	0.42	0.44	0.23	0.19	0.04	0.01
Convexa	RI	n/2	0.60	0.39	0.41	0.44	0.24	0.21	0.09	0.07
Convexa	SD	10	1.80	0.23	0.16	0.16	0.05	0.11	**0**	**0**
Convexa	SD	n/2	0.55	0.10	0.16	0.16	0.05	0.08	0.01	0.03
Valor Mínimo ECM			0.56	0.28	0.19	0.14	0.12	0.06	0.03	0.02

NULA										
Metodología			Cantidad de covariables							
Restricción	Penalidad	p	5	10	15	20	25	50	100	200
Monótona	RI	10	1.24	0.21	0.07	**0**	**0**	**0**	**0**	**0**
Monótona	RI	n/2	**0**	**0**	**0**	**0**	0.07	0.03	0.07	0.13
Monótona	SD	10	4.00	1.53	1.34	1.04	1.47	1.69	1.46	1.77
Monótona	SD	n/2	1.06	0.92	1.15	1.04	1.63	2.14	2.07	2.53
Cóncava	RI	10	4.02	1.26	0.89	0.36	0.82	0.54	0.45	0.66
Cóncava	RI	n/2	3.24	0.81	0.82	0.36	0.81	0.65	0.53	0.78
Cóncava	SD	10	11.07	3.90	3.47	2.33	3.47	3.74	3.55	4.19
Cóncava	SD	n/2	4.21	2.64	3.08	2.33	3.80	4.60	4.49	5.23
Valor Mínimo ECM			0.11	0.05	0.04	0.03	0.02	0.01	0.00	0.00

Tabla 2: Comparación de Γ-Norma para proceso Fourier $q = 1$.

CAUCHY										
Metodología			Cantidad de covariables							
Restricción	Penalidad	p	5	10	15	20	25	50	100	200
Monótona	RI	10	0.22	0.12	0.11	0.11	0.16	0.08	0.07	0.10
Monótona	RI	n/2	**0**	**0**	0.09	0.11	0.19	0.18	0.23	0.35
Monótona	SD	10	0.27	0.04	**0**	**0**	0.01	0.02	**0**	**0**
Monótona	SD	n/2	0.01	0.02	0.01	**0**	**0**	**0**	0.11	0.33
Convexa	RI	10	0.58	0.43	0.32	0.39	0.38	0.35	0.27	0.17
Convexa	RI	n/2	0.40	0.37	0.33	0.39	0.39	0.39	0.35	0.26
Convexa	SD	10	1.72	0.26	0.12	0.16	0.24	0.25	0.23	0.10
Convexa	SD	n/2	0.58	0.27	0.18	0.16	0.20	0.22	0.31	0.26
Valor Mínimo ECM			0.67	0.33	0.21	0.15	0.12	0.06	0.03	0.02

HOMOGRÁFICA										
Metodología			Cantidad de covariables							
Restricción	Penalidad	p	5	10	15	20	25	50	100	200
Monótona	RI	10	0.25	0.26	0.15	0.19	0.15	0.11	0.13	0.12
Monótona	RI	n/2	0.04	0.13	0.13	0.19	0.17	0.18	0.24	0.29
Monótona	SD	10	0.46	0.07	**0**	**0**	**0**	0.02	0.07	0.13
Monótona	SD	n/2	**0**	**0**	**0**	**0**	**0**	**0**	0.15	0.24
Convexa	RI	10	0.79	0.61	0.38	0.45	0.20	0.14	**0**	**0**
Convexa	RI	n/2	0.61	0.45	0.36	0.45	0.21	0.18	0.07	0.11
Convexa	SD	10	1.35	0.30	0.12	0.20	0.02	0.09	**0**	0.03
Convexa	SD	n/2	0.51	0.13	0.14	0.20	0.03	0.07	0.01	0.09
Valor Mínimo ECM			0.54	0.27	0.19	0.14	0.12	0.06	0.03	0.02

NULA										
Metodología			Cantidad de covariables							
Restricción	Penalidad	p	5	10	15	20	25	50	100	200
Monótona	RI	10	0.92	0.31	**0**	**0**	0.01	**0**	**0**	**0**
Monótona	RI	n/2	**0**	**0**	**0**	**0**	**0**	0.15	0.07	0.28
Monótona	SD	10	1.85	1.69	1.29	1.16	1.41	1.84	1.44	2.01
Monótona	SD	n/2	0.87	0.98	1.14	1.16	1.63	2.49	2.31	3.32
Cóncava	RI	10	2.80	1.55	0.79	0.18	0.40	0.48	0.32	0.64
Cóncava	RI	n/2	2.80	0.96	0.74	0.18	0.41	0.59	0.39	0.88
Cóncava	SD	10	12.52	4.34	3.37	2.38	2.96	3.88	3.42	4.97
Cóncava	SD	n/2	3.25	2.95	3.20	2.38	3.27	5.09	4.54	6.56
Valor Mínimo ECM			0.12	0.05	0.03	0.03	0.02	0.01	0.00	0.00

Tabla 3: Comparación de Γ-Norma para proceso Fourier $q = 0{,}2$.

CAUCHY										
Metodología			Cantidad de covariables							
Restricción	Penalidad	p	5	10	15	20	25	50	100	200
Monótona	RI	10	0.03	**0**	0.01	**0**	**0**	0.02	0.07	0.07
Monótona	RI	n/2	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Monótona	SD	10	2.32	1.39	0.94	0.85	0.93	0.63	0.40	0.19
Monótona	SD	n/2	1.24	1.06	0.88	0.85	1.00	0.78	0.58	0.36
Convexa	RI	10	0.84	1.01	0.45	0.52	0.86	0.67	0.48	0.34
Convexa	RI	n/2	1.01	0.97	0.46	0.52	0.83	0.62	0.40	0.26
Convexa	SD	10	8.31	3.64	2.30	2.40	2.71	2.18	1.32	0.75
Convexa	SD	n/2	1.63	2.35	2.21	2.40	2.76	2.48	1.86	1.26
Valor Mínimo ECM			0.12	0.05	0.04	0.02	0.02	0.01	0.01	0.00

HOMOGRÁFICA										
Metodología			Cantidad de covariables							
Restricción	Penalidad	p	5	10	15	20	25	50	100	200
Monótona	RI	10	**0**	**0**	**0**	**0**	0.01	0.03	0.07	0.09
Monótona	RI	n/2	0.06	0.01	0.01	**0**	**0**	**0**	**0**	0.02
Monótona	SD	10	2.67	0.91	0.83	0.72	0.68	0.39	0.17	**0**
Monótona	SD	n/2	1.34	0.76	0.75	0.72	0.73	0.54	0.38	0.19
Convexa	RI	10	0.74	0.53	0.51	0.62	0.49	0.38	0.29	0.17
Convexa	RI	n/2	0.87	0.60	0.53	0.62	0.48	0.34	0.23	0.11
Convexa	SD	10	7.37	2.66	2.34	2.29	1.94	1.29	0.76	0.26
Convexa	SD	n/2	1.74	1.63	2.24	2.29	1.98	1.54	1.27	0.70
Valor Mínimo ECM			0.13	0.06	0.04	0.02	0.02	0.01	0.01	0.01

NULA										
Metodología			Cantidad de covariables							
Restricción	Penalidad	p	5	10	15	20	25	50	100	200
Monótona	RI	10	**0**	**0**	**0**	**0**	**0**	0.05	0.05	0.14
Monótona	RI	n/2	0.13	0.03	0.01	**0**	**0**	**0**	**0**	**0**
Monótona	SD	10	1.66	1.28	1.00	1.11	1.16	1.27	1.07	1.53
Monótona	SD	n/2	1.03	1.09	0.95	1.11	1.23	1.64	1.62	2.39
Cóncava	RI	10	0.70	0.90	0.46	0.76	1.13	1.14	0.79	1.49
Cóncava	RI	n/2	0.98	0.89	0.47	0.76	1.04	1.00	0.68	1.21
Cóncava	SD	10	8.46	3.57	2.54	3.00	3.46	3.55	2.89	4.14
Cóncava	SD	n/2	1.75	2.27	2.45	3.00	3.60	4.13	4.12	6.09
Valor Mínimo ECM			0.13	0.05	0.03	0.02	0.01	0.01	0.00	0.00

Tabla 4: Comparación de Γ-Norma para proceso Ornstein-Uhlenbeck

CAUCHY										
Metodología			Cantidad de covariables							
Restricción	Penalidad	p	5	10	15	20	25	50	100	200
Monótona	RI	10	**0**	0.06	**0**	**0**	**0**	**0**	0.04	0.01
Monótona	RI	n/2	0.05	**0**	0.01	**0**	**0**	**0**	**0**	**0**
Monótona	SD	10	3.40	1.15	0.97	0.80	0.79	0.52	0.36	0.11
Monótona	SD	n/2	1.44	0.92	0.92	0.80	0.85	0.69	0.60	0.32
Convexa	RI	10	0.17	0.49	0.54	0.75	0.68	0.37	0.36	0.30
Convexa	RI	n/2	0.43	0.57	0.57	0.75	0.63	0.29	0.23	0.16
Convexa	SD	10	10.61	4.19	3.04	2.94	2.59	1.55	1.12	0.52
Convexa	SD	n/2	1.14	1.78	2.86	2.94	2.76	1.97	2.01	1.18
Valor Mínimo ECM			0.13	0.06	0.04	0.03	0.02	0.01	0.01	0.01

HOMOGRÁFICA										
Metodología			Cantidad de covariables							
Restricción	Penalidad	p	5	10	15	20	25	50	100	200
Monótona	RI	10	0.03	0.02	**0**	**0**	**0**	0.02	0.03	0.03
Monótona	RI	n/2	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Monótona	SD	10	3.07	1.04	0.95	0.97	0.65	0.53	0.29	0.04
Monótona	SD	n/2	1.36	0.82	0.89	0.97	0.72	0.65	0.55	0.30
Convexa	RI	10	0.74	0.58	0.57	0.97	0.27	0.49	0.27	0.34
Convexa	RI	n/2	1.07	0.66	0.59	0.97	0.28	0.36	0.15	0.20
Convexa	SD	10	12.43	4.50	3.53	3.17	1.78	1.57	0.85	0.51
Convexa	SD	n/2	1.81	2.12	3.19	3.17	1.93	2.19	1.71	1.32
Valor Mínimo ECM			0.11	0.06	0.03	0.02	0.03	0.01	0.01	0.01

NULA										
Metodología			Cantidad de covariables							
Restricción	Penalidad	p	5	10	15	20	25	50	100	200
Monótona	RI	10	**0**	**0**	**0**	**0**	**0**	**0**	0.04	0.03
Monótona	RI	n/2	0.05	**0**	0.01	**0**	0.01	0.03	**0**	**0**
Monótona	SD	10	3.13	1.38	1.14	1.36	1.25	1.07	1.21	1.31
Monótona	SD	n/2	1.99	0.98	1.09	1.36	1.35	1.41	1.97	2.43
Cóncava	RI	10	0.75	0.75	0.84	0.99	1.10	0.74	0.82	1.31
Cóncava	RI	n/2	1.20	0.87	0.81	0.99	1.00	0.56	0.54	0.85
Cóncava	SD	10	11.41	5.10	3.86	3.99	3.94	3.05	2.94	3.92
Cóncava	SD	n/2	2.04	2.27	3.70	3.99	4.11	3.97	4.83	6.85
Valor Mínimo ECM			0.11	0.05	0.03	0.02	0.02	0.01	0.00	0.00

Tabla 5: Comparación de Γ-Norma para proceso Punch.

CAUCHY										
Metodología			Cantidad de covariables							
Restricción	Penalidad	p	5	10	15	20	25	50	100	200
Monótona	RI	10	0.78	0.49	0.31	0.27	0.29	0.29	0.33	0.39
Monótona	RI	n/2	0.03	**0**	0.14	0.27	0.44	0.93	2.11	3.17
Monótona	SD	10	0.72	0.32	0.12	**0**	**0**	**0**	**0**	**0**
Monótona	SD	n/2	**0**	0.01	**0**	**0**	0.06	0.40	1.56	2.74
Convexa	RI	10	1.03	0.88	0.37	0.45	0.39	0.48	0.46	0.34
Convexa	RI	n/2	0.27	0.23	0.23	0.45	0.59	1.53	2.75	2.88
Convexa	SD	10	2.27	0.64	0.11	0.18	0.23	0.25	0.26	0.14
Convexa	SD	n/2	0.42	0.18	0.04	0.18	0.32	1.09	2.33	2.91
Valor Mínimo ECM			0.64	0.35	0.29	0.24	0.20	0.10	0.06	0.03

HOMOGRÁFICA										
Metodología			Cantidad de covariables							
Monótona	RI	p = 10	0.68	0.50	0.33	0.29	0.22	0.26	0.27	0.24
Monótona	RI	p = n/2	0.06	0.10	0.19	0.29	0.37	0.95	1.93	2.30
Monótona	SD	p = 10	1.16	0.23	0.06	**0**	**0**	**0**	0.08	0.14
Monótona	SD	p = n/2	**0**	**0**	**0**	**0**	0.06	0.40	1.43	1.90
Convexa	RI	p = 10	1.54	0.96	0.49	0.46	0.27	0.09	0.02	**0**
Convexa	RI	p = n/2	0.45	0.31	0.33	0.46	0.47	0.75	2.27	3.49
Convexa	SD	p = 10	2.39	0.60	0.23	0.25	0.10	0.03	**0**	0.01
Convexa	SD	p = n/2	0.42	0.10	0.13	0.25	0.24	0.49	1.80	2.39
Valor Mínimo ECM			0.52	0.31	0.27	0.22	0.20	0.12	0.08	0.05

NULA										
Metodología			Cantidad de covariables							
Restricción	Penalidad	p	5	10	15	20	25	50	100	200
Monótona	RI	10	3.42	0.76	0.21	**0**	**0**	**0**	**0**	**0**
Monótona	RI	n/2	**0**	**0**	**0**	**0**	0.35	0.65	0.99	4.93
Monótona	SD	10	5.79	2.07	1.23	0.81	1.23	1.36	0.88	1.28
Monótona	SD	n/2	1.08	0.79	0.84	0.81	1.66	3.08	3.58	10.19
Cóncava	RI	10	6.35	2.13	0.97	0.43	1.20	0.53	0.05	0.48
Cóncava	RI	n/2	3.36	0.66	0.60	0.43	1.47	2.48	2.05	8.03
Cóncava	SD	10	15.04	5.20	3.35	1.95	3.29	3.11	2.26	3.19
Cóncava	SD	n/2	4.33	2.34	2.28	1.95	4.25	7.05	8.97	18.65
Valor Mínimo ECM			0.08	0.05	0.04	0.04	0.02	0.01	0.01	0.00

Tabla 6: Comparación de la norma para proceso Fourier $q = 1$.

CAUCHY										
Metodología			Cantidad de covariables							
Restricción	Penalidad	p	5	10	15	20	25	50	100	200
Monótona	RI	10	0.48	0.32	0.19	0.20	0.22	0.22	0.22	0.32
Monótona	RI	n/2	0.01	**0**	0.10	0.20	0.33	0.70	1.21	2.26
Monótona	SD	10	0.47	0.17	0.06	**0**	**0**	**0**	**0**	**0**
Monótona	SD	n/2	**0**	**0**	**0**	**0**	0.07	0.29	0.80	1.93
Convexa	RI	10	0.83	0.57	0.26	0.36	0.34	0.33	0.26	0.19
Convexa	RI	n/2	0.31	0.30	0.20	0.36	0.46	0.73	1.37	1.42
Convexa	SD	10	2.14	0.37	0.06	0.10	0.20	0.16	0.12	0.03
Convexa	SD	n/2	0.49	0.21	0.05	0.10	0.26	0.47	1.30	1.78
Valor Mínimo ECM			0.68	0.36	0.28	0.21	0.17	0.09	0.05	0.03

HOMOGRÁFICA										
Metodología			Cantidad de covariables							
Restricción	Penalidad	p	5	10	15	20	25	50	100	200
Monótona	RI	10	0.48	0.41	0.20	0.21	0.19	0.12	0.29	0.23
Monótona	RI	n/2	0.07	0.10	0.14	0.21	0.29	0.51	1.31	1.66
Monótona	SD	10	0.67	0.18	0.02	**0**	**0**	**0**	0.22	0.27
Monótona	SD	n/2	**0**	**0**	**0**	**0**	0.08	0.16	1.00	1.37
Convexa	RI	10	0.93	0.77	0.38	0.40	0.16	0.03	**0**	**0**
Convexa	RI	n/2	0.50	0.38	0.28	0.40	0.27	0.44	0.87	1.81
Convexa	SD	10	1.34	0.52	0.16	0.20	0.01	0.02	0.06	0.10
Convexa	SD	n/2	0.43	0.13	0.12	0.20	0.08	0.25	0.74	1.72
Valor Mínimo ECM			0.56	0.31	0.26	0.21	0.18	0.11	0.06	0.04

NULA										
Metodología			Cantidad de covariables							
Restricción	Penalidad	p	5	10	15	20	25	50	100	200
Monótona	RI	10	1.90	0.70	0.05	**0**	**0**	**0**	0.14	**0**
Monótona	RI	n/2	**0**	**0**	**0**	**0**	0.07	0.61	0.99	3.58
Monótona	SD	10	2.65	2.03	1.11	0.98	1.08	1.46	1.28	1.60
Monótona	SD	n/2	0.87	0.88	0.92	0.98	1.50	3.04	4.27	9.15
Cóncava	RI	10	3.24	2.23	1.01	0.03	0.40	0.41	**0**	0.36
Cóncava	RI	n/2	2.57	0.82	0.78	0.03	0.50	1.41	0.94	5.68
Cóncava	SD	10	12.83	4.92	3.11	1.73	2.40	3.23	2.52	3.84
Cóncava	SD	n/2	2.99	2.62	2.80	1.73	2.91	6.59	7.37	16.31
Valor Mínimo ECM			0.11	0.05	0.04	0.04	0.02	0.01	0.01	0.00

Tabla 7: Comparación de la norma para proceso Fourier $q = 0{,}2$.

CAUCHY										
Metodología			Cantidad de covariables							
Restricción	Penalidad	p	5	10	15	20	25	50	100	200
Monótona	RI	10	0.12	**0**	**0**	**0**	0.14	0.60	1.35	2.11
Monótona	RI	n/2	**0**	0.09	0.24	**0**	**0**	**0**	0.03	0.03
Monótona	SD	10	8.52	2.88	1.71	1.08	1.50	1.10	1.44	1.97
Monótona	SD	n/2	1.20	0.83	1.11	1.08	2.62	5.19	7.44	14.74
Convexa	RI	10	0.99	0.63	0.24	0.09	0.64	1.06	1.11	1.60
Convexa	RI	n/2	1.68	0.95	0.35	0.09	0.49	0.37	**0**	**0**
Convexa	SD	10	26.70	5.73	3.51	2.37	2.63	2.71	1.23	1.16
Convexa	SD	n/2	2.85	2.06	2.38	2.37	3.89	8.83	10.44	17.33
Valor Mínimo ECM			5.67	4.54	4.75	4.61	2.97	2.00	1.70	1.18

HOMOGRÁFICA										
Metodología			Cantidad de covariables							
Restricción	Penalidad	p	5	10	15	20	25	50	100	200
Monótona	RI	10	0.05	**0**	**0**	**0**	0.11	0.77	1.98	2.72
Monótona	RI	n/2	**0**	0.35	0.15	**0**	**0**	0.05	0.14	0.21
Monótona	SD	10	5.45	2.56	1.91	1.37	1.33	1.22	2.09	2.36
Monótona	SD	n/2	1.16	1.06	1.19	1.37	2.34	5.57	12.20	19.00
Convexa	RI	10	0.66	0.48	0.12	0.24	0.35	0.42	1.32	1.73
Convexa	RI	n/2	1.19	0.77	0.17	0.24	0.18	**0**	**0**	**0**
Convexa	SD	10	21.60	5.42	3.28	2.29	2.46	1.40	1.31	1.01
Convexa	SD	n/2	2.43	1.86	1.93	2.29	3.46	6.07	11.63	16.30
Valor Mínimo ECM			6.84	5.11	5.11	4.11	3.77	2.46	1.56	1.00

NULA										
Metodología			Cantidad de covariables							
Restricción	Penalidad	p	5	10	15	20	25	50	100	200
Monótona	RI	10	**0**	**0**	**0**	**0**	**0**	1.08	1.57	2.72
Monótona	RI	n/2	0.06	0.31	0.12	**0**	**0**	**0**	**0**	**0**
Monótona	SD	10	6.05	3.20	1.79	1.34	1.30	2.12	2.56	4.33
Monótona	SD	n/2	1.01	1.39	1.25	1.34	2.43	9.78	13.66	33.67
Cóncava	RI	10	0.73	0.85	0.44	0.39	1.03	1.97	1.64	6.35
Cóncava	RI	n/2	1.31	1.13	0.58	0.39	0.62	0.76	0.10	0.73
Cóncava	SD	10	22.73	6.81	4.05	3.64	3.99	4.43	3.06	6.34
Cóncava	SD	n/2	2.39	2.36	2.83	3.64	6.96	14.59	22.97	55.31
Valor Mínimo ECM			6.44	4.09	4.13	3.56	2.11	1.15	0.71	0.28

Tabla 8: Comparación de la norma para proceso Ornstein-Uhlenbeck.

CAUCHY										
Metodología			Cantidad de covariables							
Restricción	Penalidad	p	5	10	15	20	25	50	100	200
Monótona	RI	10	**0**	**0**	**0**	**0**	0.07	0.20	0.63	0.39
Monótona	RI	n/2	0.18	0.17	0.05	**0**	**0**	**0**	**0**	**0**
Monótona	SD	10	4.22	1.55	1.35	1.07	0.84	0.79	0.78	0.34
Monótona	SD	n/2	1.70	1.13	1.03	1.07	1.32	1.65	1.82	1.37
Convexa	RI	10	0.30	0.86	1.63	2.13	3.00	2.95	5.35	5.80
Convexa	RI	n/2	0.70	1.19	2.01	2.13	2.25	0.86	0.83	0.43
Convexa	SD	10	18.90	12.23	10.72	11.03	11.63	6.45	5.66	4.51
Convexa	SD	n/2	1.51	2.23	6.66	11.03	18.86	28.48	46.43	31.79
Valor Mínimo ECM			10.57	5.34	3.61	2.89	2.47	1.71	1.16	0.90

HOMOGRÁFICA										
Metodología			Cantidad de covariables							
Restricción	Penalidad	p	5	10	15	20	25	50	100	200
Monótona	RI	10	**0**	**0**	**0**	**0**	0.09	0.18	0.40	0.46
Monótona	RI	n/2	0.13	0.14	0.11	**0**	**0**	**0**	**0**	**0**
Monótona	SD	10	4.27	1.51	1.41	1.17	0.81	0.77	0.45	0.43
Monótona	SD	n/2	1.62	1.02	1.06	1.17	1.37	1.40	2.03	1.46
Convexa	RI	10	0.94	0.97	1.51	2.35	1.98	3.98	5.78	6.23
Convexa	RI	n/2	1.37	1.22	2.05	2.35	1.36	1.33	0.87	0.59
Convexa	SD	10	23.52	12.55	13.42	11.61	6.64	5.82	5.93	5.32
Convexa	SD	n/2	2.14	2.45	7.88	11.61	15.70	41.51	47.51	33.54
Valor Mínimo ECM			8.74	5.37	3.59	2.70	2.68	1.62	1.11	0.87

NULA										
Metodología			Cantidad de covariables							
Restricción	Penalidad	p	5	10	15	20	25	50	100	200
Monótona	RI	10	**0**	**0**	**0**	**0**	0.03	0.21	0.61	0.85
Monótona	RI	n/2	0.16	0.27	0.11	**0**	**0**	**0**	**0**	**0**
Monótona	SD	10	3.70	2.32	1.59	2.00	1.65	1.45	1.87	2.37
Monótona	SD	n/2	2.28	1.39	1.38	2.00	2.32	2.56	4.14	6.43
Cóncava	RI	10	0.92	1.48	2.53	3.68	4.75	6.23	13.92	25.13
Cóncava	RI	n/2	1.49	1.74	2.95	3.68	3.20	1.98	2.71	3.29
Cóncava	SD	10	19.52	17.01	14.29	14.84	17.90	10.51	15.24	27.65
Cóncava	SD	n/2	2.35	3.04	10.15	14.84	32.02	53.71	84.83	164.46
Valor Mínimo ECM			9.05	4.20	2.95	1.79	1.62	0.99	0.46	0.19

Tabla 9: Comparación de la norma para proceso Punch.

$\widehat{b}_{mon} - u_0$	Cauchy			
	Cantidad covariables			
	25	50	75	100
Ecm	0.05	0.03	0.03	0.03
Mean	-0.12	-0.13	-0.13	-0.12
Min.	-0.38	-0.32	-0.29	-0.26
1st Qu.	-0.23	-0.21	-0.21	-0.20
Median	-0.16	-0.16	-0.16	-0.16
3rd Qu.	-0.07	-0.10	-0.10	-0.09
Max.	0.48	0.48	0.35	0.46

$\widehat{b}_{pca} - u_0$	Cauchy			
	Cantidad covariables			
	25	50	75	100
Ecm	0.06	0.05	0.05	0.04
Mean	0.02	-0.00	-0.01	-0.03
Min.	-0.33	-0.29	-0.30	-0.32
1st Qu.	-0.19	-0.20	-0.21	-0.20
Median	-0.05	-0.04	-0.05	-0.08
3rd Qu.	0.26	0.18	0.18	0.12
Max.	0.50	0.50	0.50	0.50

$\widehat{b}_{mon} - u_0$	Cuadrática			
	Cantidad covariables			
	25	50	75	100
Ecm	0.03	0.03	0.02	0.02
Mean	-0.09	-0.11	-0.12	-0.10
Min.	-0.30	-0.27	-0.27	-0.28
1st Qu.	-0.21	-0.20	-0.19	-0.17
Median	-0.12	-0.14	-0.14	-0.13
3rd Qu.	-0.02	-0.09	-0.07	-0.07
Max.	0.46	0.49	0.27	0.43

$\widehat{b}_{pca} - u_0$	Cuadrática			
	Cantidad covariables			
	25	50	75	100
Ecm	0.06	0.05	0.04	0.03
Mean	0.06	-0.03	-0.00	-0.03
Min.	-0.35	-0.32	-0.31	-0.30
1st Qu.	-0.17	-0.22	-0.18	-0.17
Median	0.08	-0.12	-0.06	-0.08
3rd Qu.	0.27	0.17	0.15	0.11
Max.	0.50	0.50	0.48	0.49

$\widehat{b}_{mon} - u_0$	Lineal			
	Cantidad covariables			
	25	50	75	100
Ecm	0.02	0.02	0.01	0.01
Mean	-0.04	-0.03	-0.03	-0.05
Min.	-0.25	-0.22	-0.21	-0.18
1st Qu.	-0.13	-0.11	-0.10	-0.11
Median	-0.05	-0.06	-0.05	-0.07
3rd Qu.	0.03	-0.00	-0.00	-0.01
Max.	0.41	0.38	0.36	0.31

$\widehat{b}_{pca} - u_0$	Lineal			
	Cantidad covariables			
	25	50	75	100
Ecm	0.05	0.03	0.03	0.02
Mean	0.07	0.03	0.00	-0.03
Min.	-0.26	-0.25	-0.23	-0.22
1st Qu.	-0.12	-0.12	-0.13	-0.13
Median	0.09	-0.04	-0.06	-0.08
3rd Qu.	0.26	0.18	0.14	0.04
Max.	0.50	0.50	0.38	0.43

$\widehat{b}_{mon} - u_0$	Escalón			
	Cantidad covariables			
	25	50	75	100
Ecm	0.01	0.01	0.01	0.01
Mean	0.04	0.03	0.03	0.04
Min.	-0.13	-0.13	-0.07	-0.05
1st Qu.	-0.02	-0.01	-0.01	0.00
Median	0.03	0.01	0.02	0.02
3rd Qu.	0.08	0.06	0.05	0.07
Max.	0.48	0.42	0.38	0.39

$\widehat{b}_{pca} - u_0$	Escalón			
	Cantidad covariables			
	25	50	75	100
Ecm	0.04	0.04	0.03	0.03
Mean	0.13	0.12	0.11	0.13
Min.	-0.21	-0.15	-0.13	-0.12
1st Qu.	0.00	-0.01	-0.00	0.02
Median	0.12	0.12	0.07	0.12
3rd Qu.	0.27	0.24	0.23	0.23
Max.	0.49	0.48	0.38	0.49

Tabla 10: Medidas de resumen para el error de estimación de u_0 por método de estimación y función estimada con covariables Fourier $q = 1$.

$\widehat{b}_{mon} - u_0$	Cauchy			
	Cantidad covariables			
	25	50	75	100
Ecm	0.20	0.20	0.18	0.19
Mean	-0.32	-0.32	-0.27	-0.30
Min.	-0.50	-0.50	-0.50	-0.50
1st Qu.	-0.50	-0.50	-0.50	-0.50
Median	-0.50	-0.50	-0.50	-0.50
3rd Qu.	-0.28	-0.27	-0.10	-0.11
Max.	0.49	0.49	0.49	0.48

$\widehat{b}_{pca} - u_0$	Cauchy			
	Cantidad covariables			
	25	50	75	100
Ecm	0.16	0.17	0.15	0.16
Mean	-0.22	-0.21	-0.18	-0.23
Min.	-0.50	-0.50	-0.50	-0.50
1st Qu.	-0.50	-0.50	-0.50	-0.50
Median	-0.38	-0.42	-0.30	-0.50
3rd Qu.	0.05	0.09	0.13	0.11
Max.	0.50	0.50	0.50	0.48

$\widehat{b}_{mon} - u_0$	Cuadrática			
	Cantidad covariables			
	25	50	75	100
Ecm	0.20	0.19	0.18	0.18
Mean	-0.31	-0.32	-0.33	-0.27
Min.	-0.50	-0.50	-0.50	-0.50
1st Qu.	-0.50	-0.50	-0.50	-0.50
Median	-0.50	-0.50	-0.50	-0.50
3rd Qu.	-0.25	-0.20	-0.14	-0.10
Max.	0.49	0.49	0.47	0.48

$\widehat{b}_{pca} - u_0$	Cuadrática			
	Cantidad covariables			
	25	50	75	100
Ecm	0.16	0.16	0.15	0.16
Mean	-0.18	-0.22	-0.24	-0.20
Min.	-0.50	-0.50	-0.50	-0.50
1st Qu.	-0.50	-0.50	-0.50	-0.50
Median	-0.35	-0.50	-0.50	-0.37
3rd Qu.	0.14	0.11	0.05	0.13
Max.	0.50	0.50	0.49	0.50

$\widehat{b}_{mon} - u_0$	Lineal			
	Cantidad covariables			
	25	50	75	100
Ecm	0.21	0.19	0.16	0.16
Mean	-0.29	-0.29	-0.23	-0.27
Min.	-0.50	-0.50	-0.50	-0.50
1st Qu.	-0.50	-0.50	-0.50	-0.50
Median	-0.50	-0.50	-0.41	-0.42
3rd Qu.	-0.18	-0.13	-0.02	-0.06
Max.	0.50	0.49	0.49	0.47

$\widehat{b}_{pca} - u_0$	Lineal			
	Cantidad covariables			
	25	50	75	100
Ecm	0.17	0.15	0.14	0.14
Mean	-0.20	-0.17	-0.17	-0.17
Min.	-0.50	-0.50	-0.50	-0.50
1st Qu.	-0.50	-0.50	-0.50	-0.50
Median	-0.50	-0.22	-0.24	-0.21
3rd Qu.	0.15	0.13	0.14	0.11
Max.	0.50	0.50	0.50	0.50

$\widehat{b}_{mon} - u_0$	Escalón			
	Cantidad covariables			
	25	50	75	100
Ecm	0.19	0.16	0.16	0.13
Mean	-0.22	-0.13	-0.12	-0.05
Min.	-0.50	-0.50	-0.50	-0.50
1st Qu.	-0.50	-0.50	-0.50	-0.50
Median	-0.50	-0.18	-0.20	-0.02
3rd Qu.	0.15	0.23	0.22	0.23
Max.	0.50	0.50	0.50	0.49

$\widehat{b}_{pca} - u_0$	Escalón			
	Cantidad covariables			
	25	50	75	100
Ecm	0.14	0.12	0.12	0.11
Mean	-0.09	-0.07	-0.10	-0.04
Min.	-0.50	-0.50	-0.50	-0.50
1st Qu.	-0.50	-0.50	-0.50	-0.38
Median	-0.03	-0.09	-0.10	0.00
3rd Qu.	0.24	0.26	0.17	0.19
Max.	0.50	0.50	0.50	0.50

Tabla 11: Medidas de resumen para el error de estimación de u_0 por método de estimación y función estimada con covariables Ornstein-Uhlenbeck.

CAUCHY									
		Fourier $q = 1$				Ornstein-Uhlenbeck			
Metodología		Cantidad de covariables							
Inicial	Tipo	25	50	75	100	25	50	75	100
Pca	3 pasos	-0.02	-0.04	-0.01	-0.02	0.08	0.06	0.00	-0.04
Pca	1 paso	1.46	1.88	2.07	1.96	8.47	5.84	4.49	3.84
Monótono	1 paso	0.08	0.04	0.06	0.02	0.24	0.27	0.06	0.01
Monótono	Oráculo	0.08	-0.03	-0.04	-0.08	-0.03	0.04	-0.08	-0.19
Ecmp Monótono 3 pasos		0.07	0.03	0.03	0.02	0.01	0.01	0.01	0.00

CUADRÁTICA									
		Fourier $q = 1$				Ornstein-Uhlenbeck			
Metodología		Cantidad de covariables							
Inicial	Tipo	25	50	75	100	25	50	75	100
Pca	3 pasos	-0.09	-0.03	0.02	-0.04	0.15	-0.01	0.07	0.01
Pca	1 paso	1.19	1.51	1.88	1.61	4.99	4.78	6.03	2.84
Monótono	1 paso	0.05	-0.04	0.04	0.00	0.30	0.09	0.11	0.08
Monótono	Oráculo	-0.10	-0.15	-0.05	-0.09	0.06	-0.13	-0.04	-0.10
Ecmp Monótono 3 pasos		0.08	0.04	0.03	0.02	0.01	0.01	0.01	0.01

LINEAL									
		Fourier $q = 1$				Ornstein-Uhlenbeck			
Metodología		Cantidad de covariables							
Inicial	Tipo	25	50	75	100	25	50	75	100
Pca	3 pasos	-0.05	-0.01	0.02	0.05	0.02	-0.01	-0.04	-0.05
Pca	1 paso	1.32	1.33	1.51	1.41	4.07	4.51	3.16	3.22
Monótono	1 paso	0.02	0.01	-0.03	-0.04	0.08	0.01	-0.00	0.06
Monótono	Oráculo	-0.16	-0.15	-0.14	-0.11	-0.14	-0.11	-0.14	-0.06
Ecmp Monótono 3 pasos		0.07	0.04	0.02	0.02	0.01	0.01	0.01	0.01

ESCALÓN									
		Fourier $q = 1$				Ornstein-Uhlenbeck			
Metodología		Cantidad de covariables							
Inicial	Tipo	25	50	75	100	25	50	75	100
Pca	3 pasos	0.05	0.20	0.20	0.14	-0.08	0.06	-0.02	0.04
Pca	1 paso	0.88	1.61	1.71	1.70	2.99	2.52	1.93	1.55
Monótono	1 paso	0.01	-0.05	0.03	-0.00	0.02	0.05	-0.04	0.02
Monótono	Oráculo	-0.24	-0.16	-0.14	-0.22	-0.11	-0.11	-0.14	-0.19
Ecmp Monótono 3 pasos		0.08	0.04	0.02	0.02	0.02	0.01	0.01	0.01

Tabla 12: Comparación de la Γ-Norma para estimadores en el Modelo Lineal Histórico.

12. Bibliografia

1. Apanasovich, T. V., & Goldstein, E. (2008). On prediction error in functional linear regression. Statistics & probability letters, 78(13), 1807-1810.

2. Asencio, M., Hooker, G., & Gao, H. O. (2014). Functional convolution models. Statistical Modelling, 14(4), 315-335.

3. Bernau, S. J. (1968). The square root of a positive self-adjoint operator. Journal of the Australian Mathematical Society, 8(1), 17-36.

4. Bertsekas, D. P. (2009). Convex optimization theory. Belmont: Athena Scientific.

5. Bosq, D. (2000). Linear Processes in Function Spaces: Theory and Applications (Vol. 149). Springer Science & Business Media.

6. Cai, T. T., & Hall, P. (2006). Prediction in functional linear regression. The Annals of Statistics, 34(5), 2159-2179.

7. Cardot, H., Ferraty, F., & Sarda, P. (1999). Functional linear model. Statistics & Probability Letters, 45(1), 11-22.

8. Cardot, H., Ferraty, F., & Sarda, P. (2003). Spline estimators for the functional linear model. Statistica Sinica, 571-591.

9. Cardot, H., Crambes, C., Kneip, A., & Sarda, P. (2007). Smoothing splines estimators in functional linear regression with errors-in-variables. Computational statistics & data analysis, 51(10), 4832-4848.

10. Cardot, H., Mas, A., & Sarda, P. (2007). CLT in functional linear regression models. Probability Theory and Related Fields, 138(3-4), 325-361.

11. Cardot, H., Goia, A., & Sarda, P. (2004). Testing for no effect in functional linear regression models, some computational approaches. Communications in Statistics-Simulation and Computation, 33(1), 179-199.

12. Cardot, H., Ferraty, F., Mas, A., & Sarda, P. (2003). Testing hypotheses in the functional linear model. Scandinavian Journal of Statistics, 30(1), 241-255.

13. Chebi, G. (2019) M-estimadores penalizados para regresión logística. Tesis de doctorado.

14. Clark, N. N., Gautam, M., Wayne, W. S., Lyons, D. W., Thompson, G., & Zielinska, B. (2007). Heavy-duty vehicle chassis dynamometer testing for emissions inventory, air quality modeling, source apportionment and air toxics emissions inventory. Coordinating Research Council, incorporated

15. Conway, J. B. (1990). A course in functional analysis. 1990. Graduate Texts in Mathematics.

16. Crambes, C., Kneip, A., & Sarda, P. (2009). Smoothing splines estimators for functional linear regression. The Annals of Statistics, 37(1), 35-72.

17. Cuesta-Albertos, J. A., García-Portugués, E., Febrero-Bande, M., & González-Manteiga, W. (2019). Goodness-of-fit tests for the functional linear model based on randomly projected empirical processes. The Annals of Statistics, 47(1), 439-467.

18. Cuevas, A. (2014). A partial overview of the theory of statistics with functional data. Journal of Statistical Planning and Inference, 147, 1-23.

19. De Boor, C. (1976). Splines as Linear Combinations of B-Splines. A Survey (No. MRC-TSR-1667). Wisconsin Univ Madison Mathematics Research Center.

20. Dubeau, F., & Savoie, J. (1996). Optimal error bounds for quadratic spline interpolation. Journal of mathematical analysis and applications, 198(1), 49-63.

21. Goia, A., & Vieu, P. (2016). An introduction to recent advances in high/infinite dimensional statistics.

22. González-Manteiga, W., & Martínez-Calvo, A. (2011). Bootstrap in functional linear regression. Journal of Statistical Planning and Inference, 141(1), 453-461.

23. Golitschek, M. V. (2014). On the L_∞-norm of the orthogonal projector onto splines. A short proof of A. Shadrin's theorem. Journal of Approximation Theory, 181, 30-42.

24. Grenander, U. (1981). Abstract inference (No. 04; QA276, G7.).

25. Guan, T., Lin, Z., & Cao, J. (2020). Estimating Truncated Functional Linear Models with a Nested Group Bridge Approach. Journal of Computational and Graphical Statistics, 1-9.

26. Hall, C. A., & Meyer, W. W. (1976). Optimal error bounds for cubic spline interpolation. Journal of Approximation Theory, 16(2), 105-122.

27. Hall, P., & Horowitz, J. L. (2007). Methodology and convergence rates for functional linear regression. The Annals of Statistics, 35(1), 70-91.

28. Hall, P., & Hooker, G. (2016). Truncated linear models for functional data. Journal of the Royal Statistical Society: Series B: Statistical Methodology, 637-653.

29. Horváth, L., & Kokoszka, P. (2012). Inference for functional data with applications (Vol. 200). Springer Science & Business Media.

30. James, G. M. (2002). Generalized linear models with functional predictors. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 64(3), 411-432.

31. James, G. M., Wang, J., & Zhu, J. (2009). Functional linear regression that's interpretable. The Annals of Statistics, 37(5A), 2083-2108.

32. Janssens, K. H., Deraedt, I., Schalm, O., & Veeckman, J. (1998). Composition of 15–17th century archaeological glass vessels excavated in Antwerp, Belgium. In Modern Developments and Applications in Microbeam Analysis (pp. 253-267). Springer, Vienna.

33. Malfait, N., & Ramsay, J. O. (2003). The historical functional linear model. Canadian Journal of Statistics, 31(2), 115-128.

34. Maronna, R. A., & Yohai, V. J. (2013). Robust functional linear regression based on splines. Computational Statistics & Data Analysis, 65, 46-55.

35. Meyer, M. C. (2008). Inference using shape-restricted regression splines. The Annals of Applied Statistics, 2(3), 1013-1033.

36. Meyer, M. C. (2012). Constrained penalized splines. Canadian Journal of Statistics, 40(1), 190-206.

37. Minka, T. P. (2000). Old and new matrix algebra useful for statistics. See www. stat. cmu. edu/minka/papers/matrix. html.

38. Morris, J. S. (2015). Functional regression. Annual Review of Statistics and Its Application, 2, 321-359.

39. Mortad, M. H. (2015). A criterion for the normality of unbounded operators and applications to self-adjointness. Rendiconti del Circolo Matematico di Palermo (1952-), 64(1), 149-156.

40. Müller, H. G., & Stadtmüller, U. (2005). Generalized functional linear models. the Annals of Statistics, 33(2), 774-805.

41. Popoviciu, T. (1933). Sur quelques propriétés des fonctions d'une ou de deux variables réelles (Doctoral dissertation, Institutul de Arte Grafice.ᴬrdealul).

42. Reimer, M. (1984). Error estimates for spline interpolants on equidistant grids. Numerische Mathematik, 44(3), 417-424.

43. Rencher, A. C., & Schaalje, G. B. (2008). Linear models in statistics. John Wiley & Sons.

44. Shadrin, A. Y. (2001). The L_∞-norm of the L_2-spline projector is bounded independently of the knot sequence: A proof of de Boor's conjecture. Acta Mathematica, 187(1), 59-137.

45. Utreras, F. (1983). Natural spline functions, their associated eigenvalue problem. Numerische Mathematik, 42(1), 107-117.

46. Van Der Vaart, A. W., & Wellner, J. A. (1996). Weak convergence. In Weak convergence and empirical processes (pp. 16-28). Springer, New York, NY.

47. Wang, J. L., Chiou, J. M., & Müller, H. G. (2016). Functional data analysis. Annual Review of Statistics and Its Application, 3, 257-29

48. Wouk, A. (1966). A note on square roots of positive operators. SIAM Review, 8(1), 100-102.

49. Yao, F., Müller, H. G., & Wang, J. L. (2005). Functional linear regression analysis for longitudinal data. The Annals of Statistics, 2873-2903.

50. Yuan, M., & Cai, T. T. (2010). A reproducing kernel Hilbert space approach to functional linear regression. The Annals of Statistics, 38(6), 3412-3444.