



Universidade de Vigo

Trabajo Fin de Máster

Series de tiempo de datos funcionales

Christian Álvarez Peláez

Máster en Técnicas Estadísticas

Curso 2019-2020

Propuesta de Trabajo Fin de Máster

Título en galego: Serie de tempo de datos funcionais
Título en español: Series de tiempo de datos funcionales
English title: Functional time series
Modalidad: Modalidad A
Autor: Christian Álvarez Peláez, Universidad de Santiago de Compostela
Director: Wenceslao González Manteiga, Universidad de Santiago de Compostela
<p>Breve resumen del trabajo:</p> <p>En los últimos años, debido a los avances metodológicos en la toma de datos, el concepto de "Series de tiempo" de datos escalares o de datos vectoriales se pudo extender a cuando estos tienen naturaleza funcional.</p> <p>El objetivo de este trabajo de fin de máster es revisar los siguientes puntos de interés relativos al estudio de una serie de datos funcionales. En particular:</p> <ul style="list-style-type: none"> a) Revisar la idea de modelo de regresión con respuesta escalar y covariable funcional. Con especial atención al modelo de regresión lineal. b) Revisar la idea de modelo de regresión con respuesta funcional y covariable funcional. Con especial atención de nuevo en el caso lineal. c) Series de tiempos, de datos funcionales, paramétricas o no paramétricas. d) El caso notable del modelo autorregresivo en espacios de Hilbert. e) Ilustración en datos reales o simulados.
<p>Recomendaciones: Se recomienda haber cursado los tópicos "Modelos de Regresión", "Modelos de Datos Funcionales" y "Series de tiempo" del máster de Técnicas Estadísticas.</p>

Don Wenceslao González Manteiga, catedrático de estadística de la Universidad de Santiago de Compostela, informan que el Trabajo Fin de Máster titulado

Series de tiempo de datos funcionales

fue realizado bajo su dirección por don Christian Álvarez Peláez para el Máster en Técnicas Estadísticas. Estimando que el trabajo está terminado, dan su conformidad para su presentación y defensa ante un tribunal.

En Santiago de Compostela, a 9 de julio de 2020.

El director:

El autor:

Don Wenceslao González Manteiga

Don Christian Álvarez Peláez

Índice general

Resumen	IX
Prefacio	XI
1. Conceptos de datos funcionales.	1
1.1. Introducción.	1
1.2. Análisis de datos funcionales.	2
1.2.1. Conceptos de análisis funcional.	2
1.2.2. Media y el operador covarianza.	5
1.3. Representación de datos funcionales.	7
1.3.1. Representación en base de funciones.	7
1.3.2. Componentes principales.	10
2. Modelos de regresión funcionales.	13
2.1. Modelos de regresión con respuesta escalar y covariable funcional.	13
2.1.1. Ejemplo ilustrativo.	15
2.2. Modelo de regresión con respuesta funcional y covariable funcional.	16
2.2.1. Ejemplo ilustrativo.	19
2.3. Mínimos cuadrados parciales (PLS).	20
2.3.1. Estimación de las componentes PLS.	22
2.4. Modelo de regresión funcional no paramétrico.	25
3. Series de tiempo de datos funcionales.	27
3.1. Conceptos de series de tiempo.	27
3.1.1. Los modelos AR, MA y ARMA.	28
3.1.2. Los modelos ARIMA.	30
3.1.3. Red neuronal autorregresiva.	30
3.2. Modelos de regresión paramétricas aplicados a las series de tiempo funcionales.	32
3.2.1. Regresión con respuesta escalar para series de tiempo funcionales.	32
3.2.2. Regresión con respuesta funcional para series de tiempo funcionales.	33
3.3. Modelos no paramétricos.	33
3.4. Modelo de red neuronal en series de tiempo de datos funcionales.	35

3.4.1. Estimación de los parámetros.	37
4. Modelo autorregresivo en espacios de Hilbert.	41
4.1. Modelo autorregresivo en espacios de Hilbert de orden 1.	41
4.1.1. Existencia.	42
4.1.2. Estimación.	44
4.1.3. Ejemplo ilustrativo.	46
4.2. Modelo ARH(p).	48
4.3. Modelo ARHX(p).	50
5. Aplicación a datos reales.	53
5.1. Medidas de error.	53
5.2. Aplicación de los modelos con respuesta funcional.	54
5.2.1. Aplicación del modelo de regresión paramétrico.	55
5.2.2. Aplicación del modelo de regresión no paramétrico.	56
5.2.3. Aplicación del modelo de red neuronal funcional.	57
5.2.4. Aplicación del modelo autorregresivo en espacios de Hilbert.	58
5.2.5. Comparativa mediante medidas de error.	59
5.3. Aplicación del modelo con respuesta escalar.	60
5.3.1. Modelo paramétrico penalizado.	61
5.3.2. Modelo de mínimos cuadrados parciales.	62
5.3.3. Comparativa mediante medidas de error.	64
6. Conclusiones y futuros trabajos.	65
A. Demostraciones importantes.	67
A.1. Demostración Lema 4.1.	67
A.2. Demostración teorema 4.1.	68
B. Estudio detallado de los datos.	71
C. Código.	73
C.1. Código modelo paramétrico.	73
C.2. código modelo no paramétrico.	74
C.3. Código del modelo red neuronal.	76
C.4. Código modelo ARH(1).	79
C.4.1. Función ARH(1).	79
C.4.2. Función Nucleophi.	80
Bibliografía	83

Resumen

Resumen en español

En este trabajo se estudian algunos de los modelos más utilizados en el análisis de datos funcionales, más concretamente en las series de tiempo de datos funcionales, donde existe algún tipo de dependencia entre las observaciones.

Se discutirán dos tipos de enfoques, un enfoque paramétrico y otro no paramétrico. Aparte nos centraremos en dos modelos distintos, los modelos de regresión y los modelos autorregresivos en espacios de Hilbert. Añadido a estos dos modelos se presenta un nuevo modelo derivado de la idea de red neuronal clásica.

Finalmente, se estudian estas distintas metodologías, aplicadas a series de tiempo de datos reales, donde se compara el comportamiento de cada una de ellas para determinar cuál es la que obtiene mejores resultados.

English abstract

The study of high-frequency data supports different methodologies. One of them is the study of functional data, where each of the observations is treated as a continuous function defined over a certain functional space.

In this work some of the models most used in the analysis of functional data are studied, more specifically in the functional time series analysis, where there is some kind of dependency between the observations.

Two types of approaches will be discussed, one parametric and one non-parametric. Besides we will focus on two different models, the regression models and the autoregressive models in Hilbert spaces. Added to these two models, a new model derived from the classical neural network idea is presented.

Finally, these different methodologies are studied, applied to time series of real data, where the behavior of each of them is compared to determine which one obtains the best results.

Prefacio

El estudio de los datos funcionales es una rama de la estadística que lleva teniendo una considerable relevancia en los últimos veinte años, pero es ahora cuando cobran una mayor importancia en una sociedad en la que el volumen de información ha crecido enormemente. Nos encontramos ante una situación en la que toda esa información requiere un análisis distinto a los métodos multivariantes clásicos. El aumento de toda esta información requiere metodologías más complejas siendo una de ellas las series de tiempo de datos funcionales.

Las series de tiempo funcionales (FTS) tratan de estudiar secuencias temporales de observaciones funcionales. Las FTS pueden ser originadas a partir de un proceso continuo en el tiempo que se divide en segmentos de igual longitud, obteniendo así una secuencia de datos funcionales. Un ejemplo de este tipo de proceso sería la evolución del precio de una acción de una empresa del Ibex 35 en función del tiempo, que se divide en segmentos diarios, obteniendo así una secuencia de precios diaria donde cada observación es una curva que describe el comportamiento en el precio a lo largo de un día dado. La imagen 1 ilustra esta transformación donde tenemos un proceso continuo que hemos dividido en intervalos de igual longitud.

Uno de los primeros planteamientos de este problema fue introducido por Bosq (2000)[5] y surge como una extensión de los modelos de series de tiempo clásicos al caso funcional. Bosq se centró en estudiar un modelo autorregresivo en el que sus elemen-

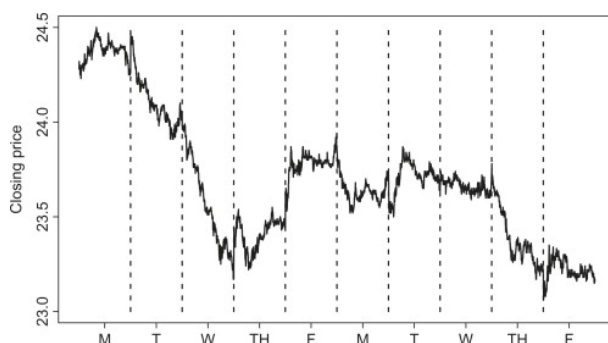


Figura 1: División en una serie de tiempo funcional

tos pertenecen a un espacio de Hilbert. Este modelo se le dio el nombre de modelos autorregresivos en espacios de Hilbert (ARH). Posteriormente surgieron otro tipo de enfoques que se centraban más en el uso de modelos de regresión funcionales. Surgieron dos corrientes distintas, la primera buscaba tratar el problema desde un enfoque enteramente paramétrico. Esta metodología se basa en las ideas propuestas por Ramsay et al.(2005)[18] . La otra corriente trataba de estudiar el problema desde una visión no paramétrica. Esta segunda visión se centraba en los trabajos de Ferraty et al.(2006)[12].

Las series de tiempo funcionales tienen numerosas aplicaciones en distintos sectores, aunque sobre todos ellos destaca uno, el de predicción en los mercados eléctricos. Algunos de los problemas que tienen especial interés en el sector eléctrico y que son abordados de una manera muy eficaz, son los estudios en la demanda eléctrica o el precio de la electricidad, donde lo más habitual es utilizar un enfoque no paramétrico.

En áreas como la del sector eléctrico, ser capaz de estudiar y predecir estas diferentes series de tiempo puede proporcionar beneficios significativos lo que motiva la investigación de este tipo de metodologías. Cualquier empresa involucrada en estos mercados estará interesada en desarrollar herramientas de toma de decisiones. Por lo tanto, el poder aplicar las FTS en casos como el mercado eléctrico que nos permitan predecir el consumo de electricidad o el precio podría producir una ventaja significativa frente al uso de otras herramientas.

El objetivo principal de este trabajo de fin de máster es estudiar las distintas metodologías de las series de tiempo funcionales que puedan proporcionar predicciones de la forma más precisa cuando sean aplicadas a series de tiempo con datos reales. En muchos casos será necesario considerar efectos de variables externas, denominadas comúnmente como variables exógenas, para poder mejorar en la medida de lo posible la predicción.

Los principales objetivos que se desarrollaran en este trabajo de fin de máster vienen resumidos en los siguientes puntos:

- Realizar un estudio sobre las distintas metodologías que se manejan actualmente en las series de tiempo funcionales.
- Aplicar estas metodologías sobre distintas series de tiempo reales.
- Comparar los resultados de las distintas metodologías con el fin de determinar cuál de todas ellas es más útil para el estudio de las series de tiempo reales con las que trabajamos.

Los objetivos descritos anteriormente se desarrollan en este trabajo a lo largo de los próximos cinco capítulos.

En el primer capítulo daremos una introducción de todos aquellos conceptos relacionados con el análisis de datos funcionales y las series de tiempo que serán necesarios para el desarrollo de este trabajo. En primer lugar, se presenta los principales conceptos y herramientas en el análisis funcional. Esto incluye una definición formal de espacios de Hilbert y algunos tipos de operadores sobre estos espacios. Luego se dará una descripción de la media y la covarianza en el marco funcional, para pasar después a describir las principales técnicas de representación de datos funcionales. En la última parte de este primer capítulo se explicarán los conceptos necesarios de las series de tiempo.

En el segundo capítulo nos centraremos en dar un análisis profundo de los modelos de regresión funcional. Nos centraremos en los modelos lineales describiendo tanto los modelos con respuesta escalar y con respuesta funcional. Las ideas aquí expuestas seguirán el marco paramétrico propuesto principalmente por Ramsay et al. (2005)[18], aunque en la parte final del capítulo describiremos una metodología distinta conocida como mínimos cuadrados parciales (PLS).

En el tercer capítulo comenzaremos a estudiar las series de tiempo de datos funcionales. En la primera parte únicamente extenderemos de forma breve las ideas expuestas en el capítulo dos a las series de tiempo. Pasaremos luego a dar una descripción enteramente no paramétrica, utilizando el estimador de Nadaraya-Watson para el análisis de las series de tiempo. En una última parte se estudiará una metodología alternativa propuesta basada en la idea de las redes neuronales.

El cuarto capítulo se enfocará en estudiar las series de tiempo funcionales desde el punto de vista de los modelos autorregresivos en los espacios de Hilbert (ARH). Nos centraremos en cuestiones como la existencia y la estimación del modelo más simple el ARH(1), para pasar a estudiar de una forma más superficial el caso de orden p . Por último daremos una extensión natural de los modelos ARH(p) añadiendo variables exógenas ARHX(p).

El quinto y último capítulo se centrará en analizar todas las metodologías descritas a lo largo del trabajo aplicándolas a un set de datos real. El set de datos en concreto es el ratio entre el precio del dólar y el euro. Realizaremos un primer enfoque centrándose en la respuesta funcional y un segundo en la respuesta escalar. Como último punto compararemos los distintos métodos y comentaremos las conclusiones al respecto.

Capítulo 1

Conceptos de datos funcionales.

En este primer capítulo se proporcionarán todas las herramientas necesarias para la comprensión del trabajo de fin de máster. Daremos una descripción cualitativa sobre que son los datos funcionales, estudiaremos los fundamentos del análisis funcional más relevantes.

1.1. Introducción.

El análisis de datos funcionales es un campo muy reciente de la estadística que sea ha ido desarrollando en los últimos veinte años, ganando relevancia a medida que han aumentado las situaciones en las que el análisis presenta datos de alta frecuencia. Se centra en el estudio de las observaciones X_1, \dots, X_n vistas como funciones definidas sobre un intervalo T . Un ejemplo clásico es el set de datos Canadian Weather (figura 1.1) donde muestra la temperaturas recogidas por 35 estaciones climatológicas en distintos puntos de Canadá. La potencia que tienen los datos funcionales es clara, poder tratar todos aquellos datos que tienen una estructura continua como pueden ser las temperaturas, señales o incluso imágenes como único ente y no como datos separados.

La principal dificultad que presentan los datos funcionales es el extrapolar las ideas de la estadística clásica a este nuevo marco. Algunas como la media o la correlación han de ser tratadas con más cuidado, otras como la función de densidad no son aplicables en el contexto de los datos funcionales. A diferencia de lo que ocurre en el caso clásico, en nuestro estudio los espacios de Hilbert serán pilar fundamental. Es por este motivo que tendrán una gran importancia ciertos operadores, generalmente operadores integrales. Otra de las grandes diferencias entre el análisis estadístico clásico y el funcional es que en este último la forma utilizada para representar nuestros datos ha de ser elegida de forma cuidadosa. Todas estas cuestiones serán discutidas en los siguientes apartados.

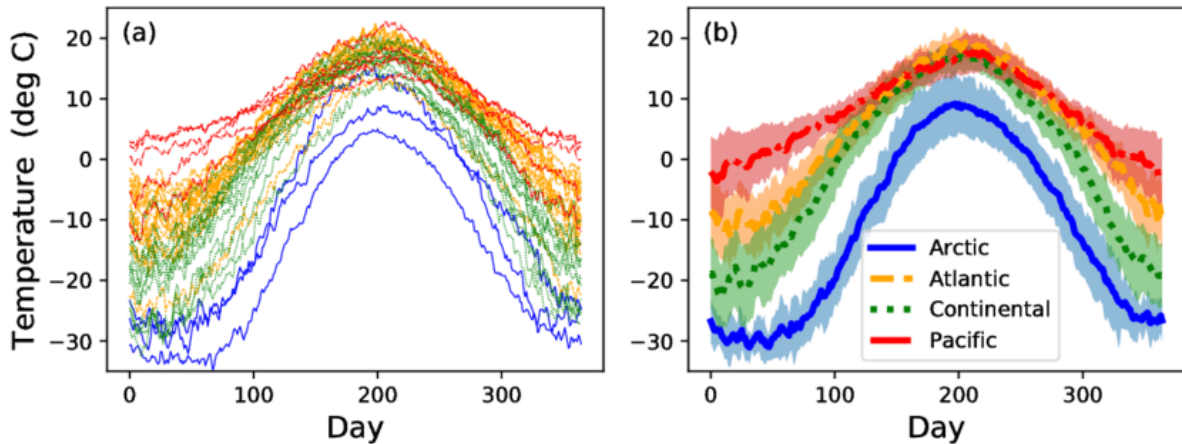


Figura 1.1: Temperaturas recogidas por 35 estaciones climatológicas en distintos puntos de Canadá

1.2. Análisis de datos funcionales.

A lo largo de esta sección daremos una descripción de las principales herramientas y metodologías necesarias para el análisis de datos funcionales. Al tratarse nuestros datos de funciones estos son infinito dimensionales, por lo tanto, las técnicas desarrolladas para la estadística multivariante no son aplicables en este caso. Comenzaremos dando una breve revisión a los conceptos del análisis funcional que serán claves para el trabajo, para luego pasar a dar una descripción sobre la media en el caso funcional y el operador de covarianza, finalizando la sección con las distintas formas de representación de los datos funcionales.

1.2.1. Conceptos de análisis funcional.

Comenzaremos explicando los conceptos del análisis funcional que serán de vital importancia para el desarrollo del trabajo. Antes de comenzar en si con la descripción de todas estas ideas daremos una definición de lo que es el pilar de nuestro análisis, la variable aleatoria funcional.

DEFINICIÓN 1.1. Diremos que la variable aleatoria X definida sobre el espacio de probabilidad (Ω, \mathbf{A}, P) es una variable aleatoria funcional si toma valores en un espacio funcional \mathcal{F} .

La definición anterior no es más que una generalización de la idea de variable aleatoria clásica al caso funcional. El marco de trabajo es muy amplio, aunque en nuestro estudio nos centraremos exclusivamente en el espacio funcional L^2 .

El espacio de L^2 es un espacio de Hilbert separable¹ que tiene como producto interno

$$\langle x, y \rangle = \int_{\Omega} x(t)y(t)dt$$

donde $x, y \in L^2$.

Dado que nuestro estudio se centra en los espacios de Hilbert, tendrá especial importancia los operadores lineales. Denotaremos por \mathcal{L} el espacio de operadores lineales en H con la norma

$$\|\Psi\|_{\mathcal{L}} = \sup\{\|\Psi(x)\| : \|x\| \leq 1\}$$

Diremos que un operador $\Psi \in \mathcal{L}$ es compacto si admite la representación

$$\Psi(x) = \sum_{j=1}^{\infty} \lambda_j \langle x, \nu_j \rangle f_j \quad x \in L^2$$

donde $\{\nu_j\}, \{f_j\}$ son bases ortonormales y $\{\lambda_j\}$ es una secuencia real que converge a 0. Esta representación recibe el nombre de descomposición en valores singulares.

En caso de que el operador admita la descomposición antes mencionada y cumpla que $\sum_{i=1}^{\infty} \lambda_i^2 < \infty$ es denominado operador de *Hilbert – Schmidt*. El espacio de operadores de Hilbert-Schmidt \mathcal{S} es un espacio de Hilbert separable con producto escalar

$$\langle \Psi_1, \Psi_2 \rangle_{\mathcal{S}} = \sum_{i=1}^{\infty} \langle \Psi_1(e_i), \Psi_2(e_i) \rangle$$

Una propiedad importante que tienen los operadores de Hilbert-Schmidt es que siempre cumplen que

¹Un espacio de Hilbert es separable si $\forall x \in H$ puede ser escrito como

$$x = \sum_k \langle e_k, x \rangle e_k$$

siendo $\{e_k\}$ una base ortonormal numerable de H .

$$\|\Psi\|_{\mathcal{L}} \leq \|\Psi\|_{\mathcal{S}}$$

Siendo $\|\Psi\|_{\mathcal{S}}^2 = \sum_{i=1}^{\infty} \lambda_i^2$.

Un operador $\Psi \in \mathcal{S}$ se dice que es simétrico si cumple

$$\langle \Psi(x), y \rangle = \langle x, \Psi(y) \rangle \quad x, y \in L^2$$

y definido positivo si

$$\langle \Psi(x), x \rangle \geq 0 \quad x \in L^2$$

Si tenemos un operador $\Psi \in \mathcal{S}$ simétrico y definido positivo entonces admite la descomposición

$$\Psi(x) = \sum_{j=1}^{\infty} \lambda_j \langle x, \nu_j \rangle \nu_j \quad x \in L^2 \quad (1.1)$$

donde ν_j son las autofunciones de Ψ .

Una clase relevante de operadores en L^2 son los operadores integrales

$$\Psi[x](t) = \int_{\Omega} \psi(t, s)x(s)ds \quad x \in L^2$$

Este tipo de operadores surgirán de forma natural en futuras secciones cuando estudiemos los modelos de regresión funcional y los autorregresivos en espacios de Hilbert. Los operadores integrales serán operadores de Hilbert-Schmidt si cumplen que

$$\int \int \psi^2(t, s) dt ds < \infty$$

En caso de que ψ sea simétrico y definido positivo admite la representación 1.1

$$\psi(t, s) = \sum_{j=1}^{\infty} \lambda_j \langle x, \nu_j(t) \rangle \nu_j(s) \quad x \in L^2 \quad (1.2)$$

1.2.2. Media y el operador covarianza.

Como se mencionó en la introducción de la sección a la hora de trabajar con datos funcionales es necesario replantearse algunos conceptos de la estadística clásica. Daremos ahora una descripción de dos conceptos que serán fundamentales en nuestro trabajo, la media funcional y el operador de covarianza.

Comenzaremos dando una descripción sobre la media funcional, para ello nos replantaremos la forma clásica de entender la media. Una forma alternativa de entender la media es pensando en ella como el centro de gravedad de los datos. De una forma genérica si estamos trabajando sobre un espacio funcional \mathcal{F} diremos que la media de una muestra $S_n = \{X_i\}_{i=1}^n$ con X_i iid es

$$\mu = E[X] = \min_{\mu \in \mathcal{F}} \sum_{X \in S} d(X, \mu)^2 \quad (1.3)$$

Es decir, la media vendrá dada como la función que minimiza la suma de las distancias al cuadrado.

DEFINICIÓN 1.2. Sea X un elemento aleatorio de L^2 diremos que es cuadrado integrable si

$$E\|X\| = E \left[\int X^2(t) dt \right]^{1/2} < \infty$$

Si X es integrable entonces existe una función $\mu \in L^2$ tal que $\mu(t) = E[X(t)]$. La esperanza tiene la propiedad de que conmuta con los operadores $\Psi \in \mathcal{L}$, $E\Psi(X) = \Psi(EX)$.

El otro objeto de estudio de esta sección es el operador de covarianzas. Si nuestra función aleatoria cumple que es cuadrado integrable entonces el operador de la covarianza queda definido como

$$C(x) = E[\langle X - \mu, x \rangle (X - \mu)] \quad x \in L^2$$

Siendo μ la media de X .

El operador de la covarianza es simétrico y definido positivo. Otra forma de escribir el operador de la covarianza es en forma integral

$$C(x)(t) = \int c(t, s)x(s)ds \quad c(t, s) = E[X(t)X(s)]$$

donde $c(t, t)$ recibe el nombre de función de covarianza.

El punto de partida es encontrar un estimador de la media adecuado. Sin embargo, el visualizar la media como el centro de gravedad de los datos implica una cuestión, que distancia debemos tomar. Es aquí donde se presenta una de las principales propiedades del análisis de datos funcionales, poder elegir sobre qué espacio funcional trabajamos y por tanto que distancia tenemos definida. Esto nos aporta flexibilidad en el análisis, pero también complejidad. Como se mencionó al principio de la sección el estudio se centra exclusivamente en el espacio L^2 .

Utilizando la definición 1.3 podemos determinar un estimador de la media. Para ello utilizaremos los siguientes desarrollos en serie $X_i = \sum_j c_{ij}\Psi_j$ y $\mu = \sum_j u_j\Psi_j$

$$\begin{aligned} \min_{\mu} \sum_j d(X, \mu)^2 &= \min_{\mu} \sum_{i=1}^n \langle X_i - \mu, X_i - \mu \rangle = \min_{\mu} \sum_{i=1}^n \langle \sum_j (c_{ij} - u_j)\Psi_j, \sum_l (c_{il} - u_l)\Psi_l \rangle \\ &= \min_{\mu} \sum_{i=1}^n (\vec{c}_i - \vec{u})^t \langle \Psi_j, \Psi_l \rangle (\vec{c}_i - \vec{u}) \end{aligned}$$

Esta última expresión es una forma cuadrática con matriz definida positiva, que toma como mínimo $\hat{u}_j = \frac{1}{n} \sum_{i=1}^n c_{ij}$

$$\hat{\mu}(t) = \sum_j \hat{u}_j \Psi_j = \frac{1}{n} \sum_{i=1}^n \sum_j c_{ij} \Psi_j = \frac{1}{n} \sum_{i=1}^n X_i(t)$$

que toma una forma análoga a la media clásica. Cabe destacar que esta forma es exclusiva de trabajar con el espacio L^2 , no podremos utilizar este estimador en otros

espacios.

A partir de esta estimación podemos determinar el estimador de la función de covarianza como

$$\widehat{c}(t, s) = \frac{1}{n} \sum_{i=1}^n (X_i(t) - \widehat{\mu}(t))(X_i(s) - \widehat{\mu}(s))$$

y el operador de la covarianza

$$\widehat{C} = \frac{1}{n} \sum_{i=1}^n \langle X_i - \widehat{\mu}, \cdot \rangle (X_i - \widehat{\mu})$$

1.3. Representación de datos funcionales.

La idea de trabajar con datos funcionales reside en tratar a toda una secuencia de datos observados como un único ente. Sin embargo, surge un problema y es que las observaciones son discretas. Estas observaciones podrán tener una mayor o menor densidad, pero nunca llegarán a ser continuas como lo es una función. Es por este motivo que la representación en base es una herramienta imprescindible en el análisis de datos funcionales. Existen una gran variedad de metodologías en la construcción de bases para datos funcionales, nosotros nos centraremos en las tres más comunes, base de fourier, B-Splines y componentes principales.

1.3.1. Representación en base de funciones.

Una de las metodologías más comunes en la representación de datos funcionales es usar bases de funciones. La metodología es simple, representar la función $x(t)$ como una expansión de la forma

$$x(t) = \sum_{i=1}^K c_i \phi_i(t)$$

donde $\phi(t)$ son los elementos de una base de funciones dada y c_i son los coeficientes del ajuste. El principal objetivo de la representación con datos funcionales es determinar el valor de los coeficientes c_i .

Una de las principales características que tiene esta representación es que el desarrollo está truncado, solo alcanza los primeros K elementos de la base. El valor de K es un factor importante puesto que cuanto mayor sea más tenderá a interpolar nuestra función. Esto implica que el uso de K determinado debe ser tratado con cuidado y se deberá elegir de forma acorde a nuestro set de datos.

La base de Fourier

La base de Fourier es posiblemente la base más conocida ya que es ampliamente utilizada en ámbitos como la ingeniería o la física. La base de Fourier está formada por los siguientes elementos

$$B = \{1, \text{sen}(\omega t), \text{cos}(\omega t), \dots, \text{sen}(n\omega t), \text{cos}(n\omega t)\}$$

por lo que nuestra estimación de la función $x(t)$ vendrá dada como

$$\hat{x}(t) = \hat{c}_0 + \hat{c}_1 \text{sen}(\omega t) + \hat{c}_2 \text{cos}(\omega t) + \hat{c}_3 \text{sen}(2\omega t) + \hat{c}_4 \text{cos}(2\omega t) + \dots$$

La forma es idéntica a la del desarrollo en serie de Fourier, la única diferencia surge en que los coeficientes c_i no pueden ser estimados con la metodología clásica del desarrollo en serie puesto que no conocemos $x(t)$.

Las bases de Fourier tienen la principal ventaja de ser bien conocidas, se pueden aplicar en la mayoría de los casos y son muy útiles cuando tenemos datos que reflejan cierta periodicidad, como pueden ser datos de temperatura. Sin embargo, tienden a ser muy suaves lo que implica que no son apropiadas para ajustar datos que tomen valores abruptos.

La base de B-splines.

Otro tipo de bases muy utilizadas son las bases de splines, en concreto la base de B-splines. Antes de dar una descripción de la base de B-splines explicaremos brevemente que es un spline.

Un spline no es más que una función construida por secciones, donde en cada sección se utiliza un polinomio de un determinado grado. En la figura 1.2 se muestra el uso de los splines para aproximar la función $\text{seno}(x)$. Al número de secciones se le suele designar como L . Para poder determinar una función spline hemos de tener un número de valores que sea superior a $L + \text{grado} + 1$ donde grado es el grado del polinomio utilizado en cada intervalo. Al número $L + \text{grado} + 1$ lo llamamos grados de libertad de la familia de splines definida sobre el conjunto de nodos. Se dice orden del spline al número $\text{grado} + 1$.

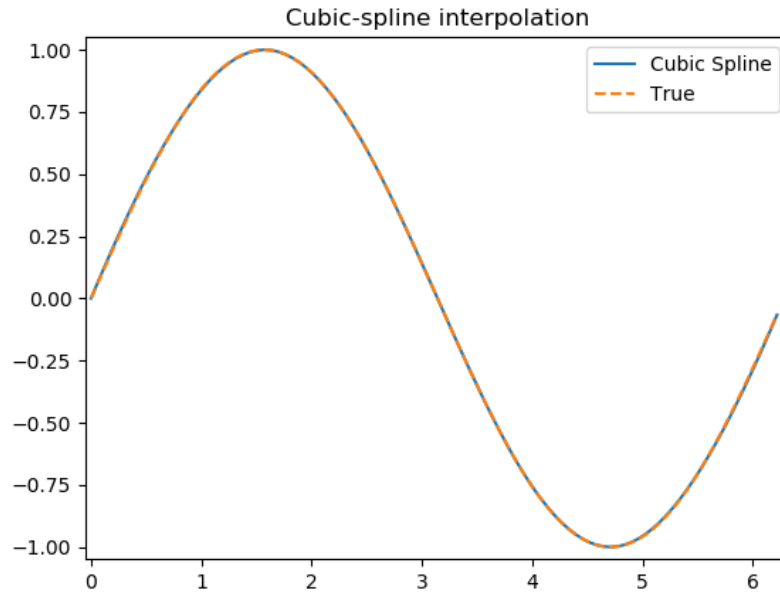


Figura 1.2: La gráfica ilustra la función $\text{seno}(x)$ y la aproximación por splines cúbicos.

La cuestión que se nos plantea ahora es determinar cómo podemos utilizar las funciones spline para construir nuestros datos funcionales. El conjunto de funciones splines es un espacio vectorial que tiene como dimensión los grados de libertad. Esto implica que se pueden utilizar diversas bases para la representación. En nuestro análisis utilizaremos la más común de todas, la base $B - \text{Splines}$ (figura 1.3) que fue propuesta por Boor (1978)[4].

Podemos por tanto representar cualquier función spline $S(t)$ como

$$S(t) = \sum_{i=1}^{L+\text{orden}} c_i B_i(t, \tau)$$

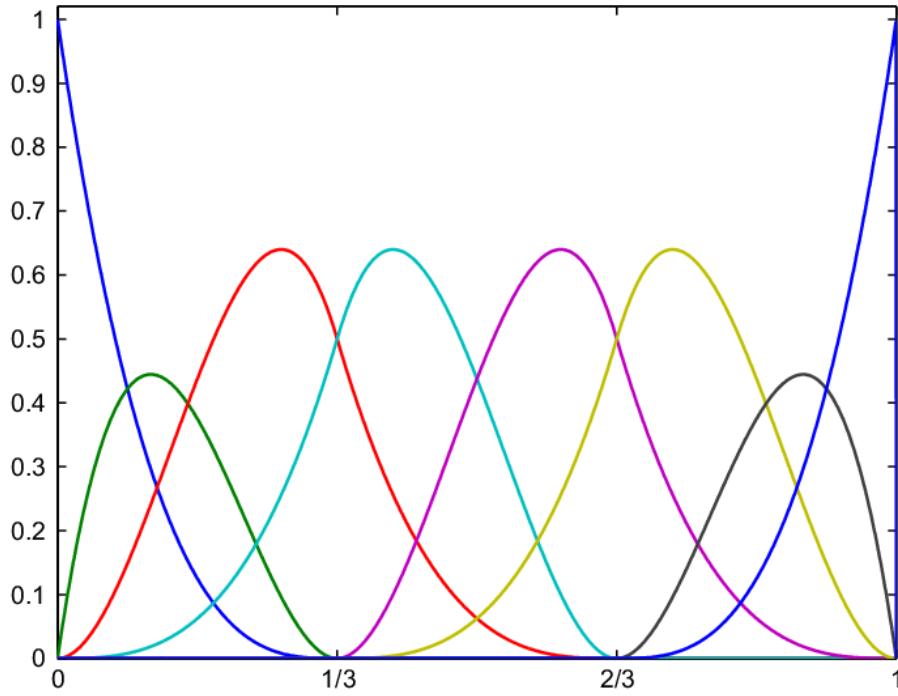


Figura 1.3: La gráfica ilustra las funciones de la base de B-spline cúbicas.

donde B_i es la i -ésima función de la base B-splines que viene definida sobre el conjunto de nodos τ en el tiempo t , mientras que c_i son los coeficientes correspondientes a la función Spline en dicha base.

1.3.2. Componentes principales.

La siguiente metodología de estudio es posiblemente una de las más importantes en todo el análisis de datos funcionales, las componentes principales de datos funcionales (FPC). Al igual que ocurría con el caso clásico las FPC nos permiten reducir la dimensión de datos funcionales, que tienen dimensión infinita a dimensión finita de una manera óptima.

Las ideas del análisis de componentes principales de la estadística multivariante se pueden extender fácilmente a los espacios de Hilbert separables. Supongamos que $\Psi(x)$ es un operador de Hilbert-Schmidt simétrico definido positivo en L^2 . Este operador admite la descomposición espectral 1.1. Tenemos por tanto el siguiente problema de maximización

$$\langle \Psi(x), x \rangle = \left\langle \sum_{j=1}^{\infty} \lambda_j \langle x, \nu_j \rangle \nu_j, x \right\rangle = \sum_{j=1}^{\infty} \lambda_j \langle x, \nu_j \rangle^2$$

sujeito a $\|x\| = 1$. La identidad de Parseval nos dice que $\|x\|^2 = \sum_{j=1}^{\infty} \langle x, \nu_j \rangle^2$ por

lo que tenemos $\sum_{j=1}^{\infty} \langle x, \nu_j \rangle^2 = 1$. Tomaremos la primera componente principal como aquella función que cumple $\langle x, \nu_1 \rangle^2 = 1$ y $\langle x, \nu_j \rangle = 0 \quad j > 1$. Tenemos que la función que maximiza $\langle \Psi(x), x \rangle$ es ν_1 , que toma como máximo el autovalor λ_1 . Para determinar la segunda componente el proceso es exactamente idéntico al anterior añadiendo la condición $\langle x, \nu_1 \rangle = 0$, de esta forma la segunda componente principal será ν_2 y el valor máximo será el autovalor λ_2 . Repetimos este procedimiento para determinar el resto de componentes principales. Podemos generalizar este procedimiento en el siguiente teorema.

TEOREMA 1.1. Si tenemos un operador de Hilbert-Schmidt Ψ que es simétrico y semidefinido positivo con autivalores λ_i y autofunciones ν_i , entonces

$$\sup\{\langle \Psi(x), x \rangle : \|x\| = 1, \langle x, \nu_i \rangle = 0, 1 \leq i \leq j-1\} = \lambda_j$$

donde el supremo es alcanzado si $x = \nu_i$

La base de componentes principales.

Nuestro objetivo ahora es determinar una base ortonormal η_i que nos minimice

$$\widehat{S}^2 = \sum_{j=1}^N \|x_j - \sum_{i=1}^p \langle x_j, \eta_i \rangle \eta_i\|^2$$

El valor p será el número de bases que deseamos utilizar en nuestra representación, que ha de ser siempre menor que nuestro número de datos. El usar un mayor número de elementos no requiere necesariamente un mejor rendimiento en nuestro análisis.

Podemos reescribir la expresión anterior de la siguiente forma

$$\begin{aligned} \sum_{j=1}^N \|x_j - \sum_{i=1}^p \langle x_j, \eta_i \rangle \eta_i\|^2 &= \sum_{j=1}^N \|x_j\|^2 - 2 \sum_{j=1}^N \sum_{i=1}^p \langle x_j, \eta_i \rangle^2 + \sum_{j=1}^N \sum_{i=1}^p \langle x_j, \eta_i \rangle^2 \|\eta_i\|^2 \\ &= \sum_{j=1}^N \|x_j\|^2 - \sum_{j=1}^N \sum_{i=1}^p \langle x_j, \eta_i \rangle^2 \end{aligned}$$

Es decir, nuestro problema de minimización queda reducido a encontrar las autofunciones que maximicen $\sum_{j=1}^N \sum_{i=1}^p \langle x_j, \eta_i \rangle^2$, que puede ser reescrito como $\langle \widehat{C} \eta_i, \eta_i \rangle$, siendo \widehat{C} el operador de covarianza. Dado que el operador de la covarianza es un operador de Hilbert-Schmidt simétrico y definido positivo concluimos por el teorema 1.1 que la base óptima es la base de componentes principales ν_i .

Capítulo 2

Modelos de regresión funcionales.

A lo largo de este capítulo nos centraremos en dar una descripción de los principales fundamentos de los modelos de regresión funcional. En la primera sección analizaremos el caso notable de los modelos de regresión con respuesta escalar y covariable funcional paramétrico, estudiando brevemente su aplicabilidad a un caso simulado. En una segunda sección nos centraremos en los modelos de regresión con respuesta funcional y covariable funcional paramétrico, y al igual que para el caso escalar estudiaremos su aplicabilidad con un ejemplo simulado. En la siguiente sección nos centraremos en los mínimos cuadrados parciales (PLS) que consiste en describir la regresión funcional paramétrica con una metodología más enfocada en la predicción. Por último, describiremos brevemente el modelo de regresión funcional no paramétrico.

2.1. Modelos de regresión con respuesta escalar y covariable funcional.

Es interesante para muchas situaciones determinar la relación que existe entre una variable funcional y una variable escalar, es por este motivo que la regresión con respuesta escalar y covariable funcional tiene gran importancia. Para contextualizar un poco la aplicación y el interés de este tipo de regresión ejemplificaremos utilizando el set de datos Candian Weather del capítulo anterior (imagen 1.1). Imaginemos que estamos interesados en estudiar si las precipitaciones totales a lo largo del año son dependientes del registro de temperaturas de cada una de las estaciones. En este caso nuestras temperaturas recogidas a lo largo de todo el año son un dato funcional, mientras que las precipitaciones totales son un valor escalar.

En nuestro análisis del modelo de regresión con respuesta escalar supondremos que nuestra variable funcional $X \in L^2$ y que $y \in \mathbb{R}$. También asumiremos que $\mathbb{E}[X(t)] = 0$ y $\mathbb{E}[y] = 0$. Esta suposición sobre la media a priori puede parecer muy restrictiva, sin embargo, en el caso de que tengamos una variable $\mathbb{E}[X(t)] = \mu(t)$ realizaremos nuestro

análisis con $X'(t) = X(t) - \hat{\mu}(t)$.

El modelo de regresión lineal viene definido como:

$$y_i = \langle X_i, \beta \rangle + \epsilon = \int X(t)\beta(t)dt + \epsilon$$

Con $\beta \in L^2$ y ϵ el término del error.

El objetivo por tanto será conseguir determinar un buen estimador de la función β . Para ello aproximaremos X_i y β usando un número finito de elementos de la base.

$$X_i(t) = \sum_j^{A_x} c_{ij}\psi_j(t), \quad \beta(t) = \sum_j^{A_\beta} b_j\theta_j(t)$$

Al utilizar un número finito de elementos de la base nos permite poder describir el problema en forma matricial.

$$X = C\Psi, \quad \beta = \Theta^t b$$

Tenemos por tanto que

$$y = \langle X, \beta \rangle + \epsilon \simeq C\Psi\Theta^t b + \epsilon = Zb + \epsilon$$

Hemos reducido el problema al modelo lineal clásico, donde el estimador por mínimos cuadrados de b es

$$b = (Z^t Z)^{-1} Z^t y$$

Y nuestro estimador \hat{y} es

$$\hat{y} = Z(Z^t Z)^{-1} Z^t y = Hy$$

Esta metodología trae consigo una dificultad, y es el poder determinar la inversa de $(Z^t Z)$. La problemática surge en el hecho de que Z es una matriz en la que su dimensión depende del número de elementos de la base elegida. Esto implica situaciones en las que el uso elevado de elementos de la base pueda provocar el no poder determinar el modelo de regresión. Para evitar este problema Ramsay et al.(2005)[18] proponen una versión penalizada de la suma residual de cuadrados

$$PENSSE_\lambda(\beta) = \sum_i^N \left[y_i - \sum_j^{A_\beta} \langle X_i, \theta_j \rangle b_j \right]^2 + \lambda \sum_{j=1}^{A_\beta} b_j^2 \int \left[\theta_j^m(s) \right]^2 ds$$

donde hemos utilizado la representación $\beta(t) = \sum_j^{A_\beta} b_j \theta_j(t)$ y el factor de penalización es proporcional a la integral del cuadrado de la derivada de orden m de β .

Este método se puede reescribir como

$$PENSSE_\lambda(b) = \|y - Zb\| + \lambda b^t R b$$

que toma como \hat{b} mínimo

$$b = (Z^t Z + \lambda R)^{-1} Z^t y$$

Donde la selección del valor del parámetro λ se realiza por validación cruzada.

2.1.1. Ejemplo ilustrativo.

Para contextualizar todas estas ideas, estudiaremos como es la estimación de nuestro parámetro β en un modelo teórico. Definiremos nuestra variable funcional $X(t)$ como una colección de procesos gaussianos y nuestra variable a estimar $\beta(t) = t + \log(t+0,1)$. El modelo por tanto viene definido como

$$y = \langle X(t), \beta(t) \rangle + \epsilon$$

donde para este ejemplo utilizaremos que los residuos sigan una distribución $N(0, 1)$.

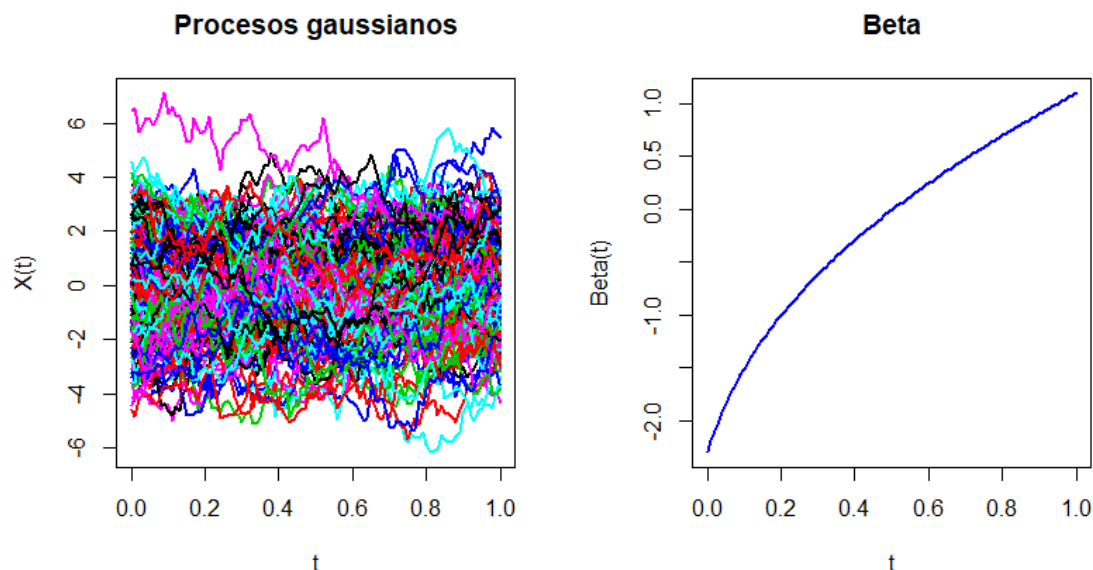


Figura 2.1: Variables funcionales $X(t)$ y $\beta(t)$

Utilizaremos la base de funciones de B-Splines para nuestra representación. En la imagen 2.2 se representa en azul el valor teórico de nuestra función *beta* y en rojo discontinua nuestra estimación. Como podemos ver la aproximación es bastante buena, aunque tiende a tener una mayor curvatura. Esto es algo que ocurre frecuentemente cuando trabajamos con las bases de funciones.

2.2. Modelo de regresión con respuesta funcional y covariable funcional.

Cuando queremos estudiar la relación existente entre dos variables y ambas son funcionales entra en juego los modelos de regresión con respuesta funcional y covariable funcional. Un ejemplo clásico de la regresión con respuesta funcional es determinar si existe alguna relación entre las temperaturas a lo largo del año en una ciudad y el consumo de luz de esta. El modelo funcional lineal toma la siguiente forma:

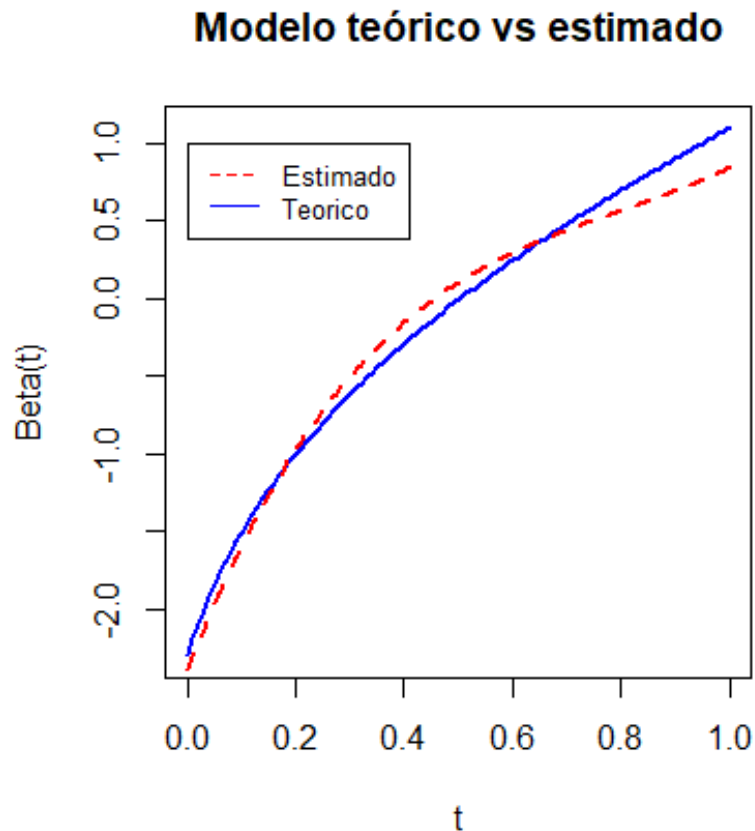


Figura 2.2: Comparación modelo teórico y estimado.

$$Y_i(t) = \alpha(t) + \int \beta(t, s)X_i(s)ds + \epsilon_i(t)$$

donde $\alpha(t)$ es una función que tendría como análogo clásico el intercepto, $\epsilon_i(t)$ son los residuos y β es el operador a estimar. A diferencia de lo que ocurría en la regresión respuesta escalar nuestros residuos tienen un carácter funcional. Esta característica nos da un grado mayor de dificultad en el análisis respecto del caso escalar. Es relativamente común que en algunos análisis no utilicen la función $\alpha(t)$. En caso de que nuestro modelo no incluya esta función α podemos describir el modelo utilizando un operador integral:

$$Y_i(t) = \Psi(X_i) + \epsilon_i(t)$$

donde

$$\Psi[x](t) = \int \beta(t, s)x(s)ds$$

Esta forma es muy similar a la que toman los modelos autorregresivos en espacios de Hilbert, como veremos más adelante.

Para ajustar este modelo de regresión la situación es algo más compleja al caso con respuesta escalar. Una forma usual de realizar el ajuste es utilizando la suma residual de cuadrados integrada:

$$LMSSE(\beta) = \int \sum_{i=1}^n \left[Y_i - \int \beta(t, s)X_i(s)ds \right]^2 dt$$

El siguiente paso es reescribir β como una doble expansión en términos de dos bases η_k y θ_l

$$\beta(s, t) = \sum_{k=1}^K \sum_{l=1}^L b_{kl} \eta_k(s) \theta_l(t) = \eta' \mathbf{B} \theta$$

Si fijamos el número de elementos base de X , tenemos:

$$Y(t) = \int X(s) \eta'(s) B \theta(t) ds + \epsilon(t) = \mathbf{X} \mathbf{B} \theta(t) + \epsilon(t)$$

donde $\mathbf{X} = \int X(s) \eta'(s) ds$ es una matriz $n \times K$. Multiplicamos en ambos lados de la igualdad por $\mathbf{X}' \int \theta(t) dt$

$$\mathbf{X}' \mathbf{X} \mathbf{B} \int \theta(t) \theta'(t) dt = \mathbf{X}' \int Y(t) \theta'(t) dt$$

Esta expresión puede ser reescrita como:

$$[J_{\theta\theta} \otimes (\mathbf{X}'\mathbf{X})]vec(\mathbf{B}) = vec\left(\mathbf{X}' \int Y(t)\theta'(t)dt\right)$$

donde $J_{\theta\theta} = \int \theta(t)\theta'(t)dt$ y \otimes es el producto de Kronecker. Si representamos la respuesta como $\mathbf{Y} = \mathbf{C}\phi$ entonces tenemos:

$$vec(\widehat{\mathbf{B}}) = [J_{\theta\theta} \otimes (\mathbf{X}'\mathbf{X})]^{-1}(J_{\phi\theta} \otimes \mathbf{X}')vec(\mathbf{C})$$

Análogamente al caso escalar podemos introducir algún tipo de penalización

$$vec(\widehat{\mathbf{B}}) = [J_{\theta\theta} \otimes (\mathbf{X}'\mathbf{X}) + \lambda_s J_{\theta\theta} \otimes \mathbf{R} + \lambda_t \mathbf{Q} \otimes J_{\eta\eta}]^{-1}(J_{\phi\theta} \otimes \mathbf{X}')vec(\mathbf{C})$$

Para el caso funcional el uso de las penalizaciones resulta aún más necesario que en el caso de respuesta escalar.

2.2.1. Ejemplo ilustrativo.

Al igual que hicimos con el caso escalar vamos a plantear un modelo teórico a estudiar. Utilizaremos un modelo con $\beta(s, t) = e^{-\frac{t^2+s^2}{2}}$ y $\alpha(t) = 0$ donde los residuos vendrán dados como una combinación lineal de funciones trigonométricas de la forma $\epsilon(t) = 0,5 * N_t(0, 1) * \text{sen}(\pi t) + 0,5 * N_t(0, 1) * \text{sen}(\pi t)$. La variable $N_t(0, 1)$ no es más que una $N(0, 1)$ para cada instante de tiempo. La variable funcional $X(t)$ la definimos exactamente igual que para el caso de respuesta escalar.

En la imagen 2.3 se muestra el resultado de nuestra estimación. Como puede apreciarse la estimación es mucho más rugosa. La causa principal de esta rugosidad es el uso de un número bajo de elementos de la base para representar nuestra variable respuesta y que en general los modelos de regresión funcional tienden a ser menos suaves que las funciones que estiman. De todas formas, la estimación de nuestra β es muy buena. A pesar de que nuestro modelo no incluía la función $\alpha(t)$ en el ajuste no hemos añadido esta restricción con el objetivo de comprobar como de buena es la estimación. En el caso ideal nuestra estimación debería ser 0 o una función que tome valores muy bajos.

En la imagen 2.4 se muestran que los valores que toma el intercepto son muy próximos a 0, con lo que podemos determinar que la estimación es en general considerable-

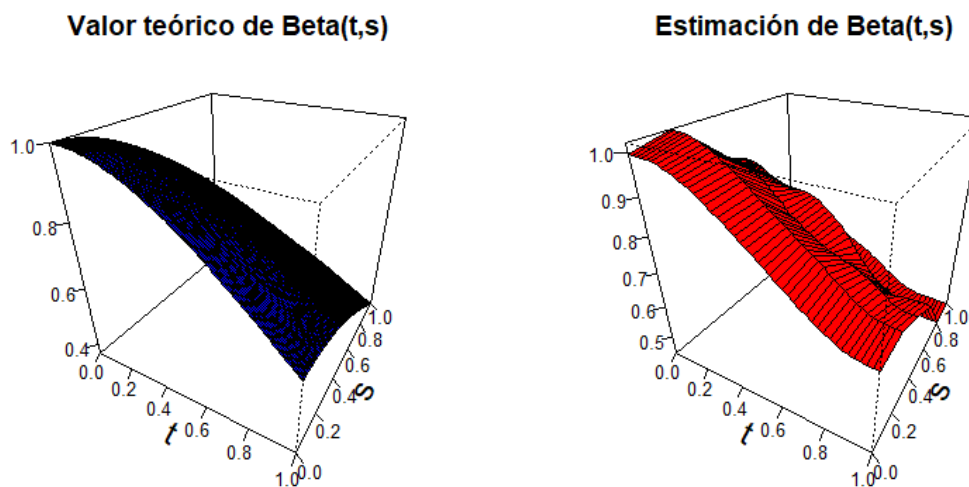


Figura 2.3: Comparación modelo teórico y estimado.

mente buena.

2.3. Mínimos cuadrados parciales (PLS).

En las dos secciones anteriores hemos estudiado los modelos de regresión con respuesta escalar y funcional. Sin embargo, la metodología para estimar los modelos ha sido minimizando la suma residual de cuadrados añadiendo una penalización. Esta metodología se le denomina mínimos cuadrados y es útil en situaciones en las que nuestros residuos presenten homocedasticidad y sean independientes, sin embargo, como se mencionó anteriormente presenta el inconveniente de que se centra en la inversión de una matriz que depende directamente del número de elementos que elegimos de la base. Esto causa que cuando usamos un número de elementos de la base elevado la inversión de la matriz no sea posible. Para solucionar esta problemática se utilizan la penalización, aunque ello implica tener que determinar los valores de dos parámetros en el caso funcional y uno en el escalar, lo que repercute en un mayor coste computacional. Existen distintas metodologías para resolver este problema, nosotros nos centraremos en los mínimos cuadrados parciales.

El propósito de los mínimos cuadrados parciales es estimar una base que tenga como objetivo conseguir la mejor predicción posible. La idea es poder descomponer nuestra variable respuesta y la covariable en términos de variables aleatorias, de forma que

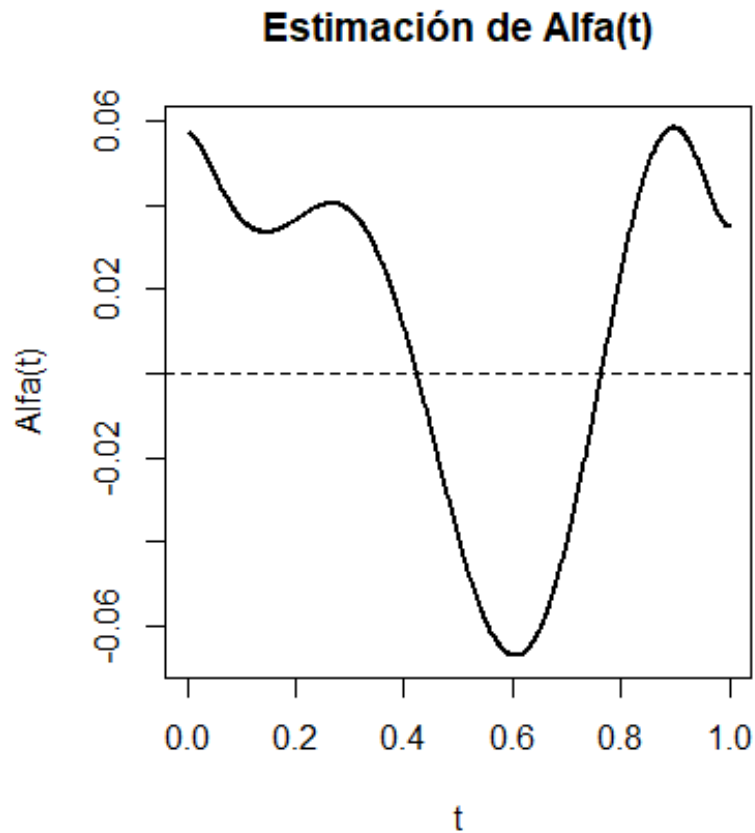


Figura 2.4: Estimación $\alpha(t)$.

maximicemos el acierto en la predicción. Al igual que en los casos anteriores asumimos que nuestras funciones aleatorias tienen media cero. Para obtener las componentes PLS funcionales maximizaremos

$$\max_{\substack{\kappa \in L^2, \|\kappa\|=1 \\ \zeta \in L^2, \|\zeta\|=1}} Cov^2 \left(\int X(s)\kappa(s)ds, \int y(t)\zeta(t)dt \right)$$

Esta maximización es exclusiva para el caso de la regresión con respuesta funcional. Para el caso de respuesta escalar la maximización es

$$\max_{\kappa \in L^2, \|\kappa\|=1} Cov^2 \left(\int X(s)\kappa(s)ds, y \right)$$

Las componentes PLS funcionales se corresponden con las autofunciones del operador de Escoufier (para una visión más profunda Preda et al. (2005)[17]). Nosotros no nos centraremos en el análisis de este operador, simplemente tomaremos los resultados de Beyaztas et al.(2020)[3] para el caso con respuesta funcional y Febrero-Bande et al.(2017)[9] para el caso con respuesta escalar.

2.3.1. Estimación de las componentes PLS.

La forma de obtener las componentes PLS es recursiva. Partimos de la definición de la primera componente PLS:

$$\eta_1 = \int \kappa_1(s)X(s)ds$$

donde la función κ_1 es una función peso que viene definida como

$$\kappa_1(s) = \frac{\int E[X(s)y(t)]dt}{\|\int E[X(s), y(t)]dt\|}$$

Una vez que hemos determinado la primera componente PLS calcularemos la segunda utilizando los residuos de los siguientes modelos de regresión:

$$\begin{aligned} X_1(s) &= X_0(s) - p_1(s)\eta_1 \\ y_1(t) &= y_0(t) - \zeta_1(t)\eta_1 \end{aligned}$$

en este caso X_0 y y_0 son nuestras variables aleatorias originales y las funciones $p_1(s) = \frac{Cov(X_0(s)\eta_1)}{Var(\eta_1)}$ y $\zeta_1 = \frac{Cov(y_0(t)\eta_1)}{Var(\eta_1)}$. Para determinar esta vez la componente η_h PLS tenemos:

$$\eta_h = \int \kappa_h(s)X_{h-1}(s)ds$$

donde la función peso viene dada por:

$$\kappa_h(s) = \frac{\int E[X_{h-1}(s)y_{h-1}(t)]dt}{\|\int E[X_{h-1}(s), y_{h-1}(t)]dt\|}$$

Antes de continuar con el análisis de las componentes PLS es conveniente enunciar el siguiente lema y teorema.

LEMA 2.1 (Lema de Mercer). Sea c una función de covarianza continua. Entonces existe una secuencia (ψ_n) de funciones continuas y una secuencia decreciente (λ_n) positiva tal que

$$\int_0^1 c(s, t)\psi_n(s)ds = \lambda_n\psi_n(t) \quad n \in \mathbb{N}$$

y

$$\int_0^1 \psi_n(s)\psi_m(s)ds = \delta_{n,m} \quad n, m \in \mathbb{N}$$

Además,

$$c(s, t) = \sum_{n=0}^{\infty} \lambda_n \psi_n(s)\psi_n(t)$$

donde la serie converge de forma uniforme, por lo tanto

$$\sum_{n=0}^{\infty} \lambda_n = \int c(t, t)dt < \infty$$

TEOREMA 2.1 (Expansión de Karhunen-Loève). Sea $X = (X_t)$ un proceso estocástico cuadrado integrable de media cero definido sobre un espacio de probabilidad Ω , con función de covarianza continua c . Entonces

$$X_t = \sum_{n=0}^{\infty} \eta_n \psi_n(t)$$

donde η_n es una secuencia real de variables aleatorias de media cero tal que

$$E(\eta_n \eta_m) = \lambda_n \delta_{n,m} \quad n, m \in \mathbb{N}$$

y donde la secuencia (λ_n, ψ_n) es definida por el lema 2.1. La serie converge de forma uniforme en Ω .

Una vez determinadas las componentes PLS podemos escribir nuestras funciones aleatorias utilizando el teorema 3.1 como:

$$\begin{aligned} X(s) &= \sum_{h=1}^{\infty} p_h(s) \eta_h \\ y(t) &= \sum_{h=1}^{\infty} \zeta_h(t) \eta_h \end{aligned}$$

Podemos reescribir η_h como $\eta_h = \langle X_{h-1}, \kappa_h \rangle = \langle X, \phi_h \rangle$ donde $\phi_h = \kappa_h - \langle p_1, \kappa_h \rangle \phi_1 - \dots - \langle p_{h-1}, \kappa_h \rangle \phi_{h-1}$ siendo $\phi_1 = \kappa_1$. Tenemos por tanto que podemos reescribir la variable $y(t)$ de la siguiente manera

$$y(t) = \sum_h^{\infty} \zeta_h(t) \langle X_{h-1}(s), \kappa_h(s) \rangle = \sum_h^{\infty} \zeta_h(t) \langle X, \phi_h(s) \rangle = \langle X, \sum_h^{\infty} \zeta_h(t) \phi_h(s) \rangle$$

Si llamamos a $\sum_h^{\infty} \zeta_h(s) \phi_h(s) = \beta(s, t)$ llegamos a la expresión del modelo de regresión con respuesta funcional

$$y(t) = \int X(s) \beta(s, t) ds$$

por lo que nuestra estimación de β queda determinada calculando las PLS.

Para el caso con respuesta escalar el procedimiento es idéntico al descrito anterior, con la diferencia de que nuestra variable respuesta es en este caso un número y no una función. La función $\kappa_h(s)$ viene definida entonces como

$$\kappa_h(s) = \frac{\text{cov}(X_{h-1}(s), y_{h-1})}{\|\text{cov}(X_{h-1}(s), y_{h-1})\|}$$

y las variables η_h y $p_h(s)$

$$p_h(s) = \frac{\text{cov}(X_{h-1}(s), \eta_{h-1})}{\text{var}(\eta_{h-1})}$$

$$\eta_h = \frac{\text{cov}(y, \eta_{h-1})}{\text{var}(\eta_{h-1})}$$

Extrapolando el mismo análisis para las funciones ϕ realizado para el caso funcional, podemos reescribir nuestra variable y de la siguiente manera

$$y = \sum_h^{\infty} \zeta_h \langle X, \phi_h(s) \rangle = \langle X, \sum_h^{\infty} \zeta_h \phi_h(s) \rangle = \langle X, \beta \rangle$$

Siendo en este caso $\beta(s) = \sum_h^{\infty} \zeta_h \phi_h(s)$.

2.4. Modelo de regresión funcional no paramétrico.

Otro enfoque completamente distinto es el uso de modelos no paramétricos para la regresión funcional. Esta metodología fue propuesta principalmente por Ferraty et al.(2006)[12]. La regresión no paramétrica intenta estimar un operador funcional, pero a diferencia de los modelos paramétricos, no se asume la necesidad de que el modelo siga una estructura lineal. De una forma general podemos definir el modelo no paramétrico como:

$$Y_i = m(X_i) + \epsilon_i \quad i = 1, \dots, n$$

La variable aleatoria Y_i puede ser funcional (definida sobre un espacio funcional \mathcal{H}) o real, mientras que la X_i es una variable aleatoria definida sobre el espacio funcional \mathcal{H}' . La función m es la función de regresión y ϵ_i un término de error. Al igual que para el caso paramétrico solamente trabajaremos con espacios de Hilbert, aunque cabe destacar que es bastante habitual el uso de espacios Banach en esta metodología.

La estimación de m es un tema crucial en particular para predecir el valor de la respuesta dada una nueva variable funcional explicativa X_{n+1} . Para proporcionar un procedimiento consistente para estimar el operador de regresión no lineal m , es común

adaptar el estimador de Nadaraya-Watson clásico al caso funcional.

$$\hat{m}(X) = \frac{\sum_{i=1}^n Y_i K(h^{-1}\|X_i - X\|)}{\sum_{i=1}^n K(h^{-1}\|X_i - X\|)}$$

La estimación no paramétrica tiene como principal ventaja su flexibilidad ya que no asume linealidad en el modelo. Sin embargo, la determinación de una ventana h puede ser un problema serio. Esta situación es ya bien conocida de la estadística no paramétrica clásica, aunque en el caso funcional es aun más necesario determinar la ventana de una forma adecuada. Una revisión profunda sobre la regresión funcional no paramétrica se puede encontrar en Ferraty et al.(2016) [10].

Capítulo 3

Series de tiempo de datos funcionales.

A lo largo de todo nuestro estudio nos hemos centrado en el estudio de datos funcionales bajo una premisa, nuestros datos son independientes y están idénticamente distribuidos. Esta suposición puede ser considerablemente restrictiva teniendo en cuenta la naturaleza y procedencia de los datos. En situaciones como las transacciones financieras, los datos son recogidos de forma secuencial a lo largo del tiempo, es razonable pensar que bajo esta situación nuestras observaciones se vean afectadas por las observaciones pasadas. Es por este motivo que se requieren herramientas adicionales para analizar los datos funcionales que tienen un comportamiento similar al de las series de tiempo. Surgen así las series de tiempo de datos funcionales, que son el objeto de estudio principal de este trabajo.

Existen distintas metodologías a la hora de estudiar las series de tiempo funcionales, aunque en nuestro estudio nos centraremos en tres puntos de vista, los métodos de regresión paramétrica, no paramétrica y los modelos autorregresivos en espacios de Hilbert. En este capítulo analizaremos los dos primeros puntos de vista, dejando los modelos autorregresivos en espacios de Hilbert para el capítulo posterior.

3.1. Conceptos de series de tiempo.

Antes de comenzar con las series de tiempo de datos funcionales presentaremos aquellos conceptos más necesarios de las series de tiempo. Comenzaremos dando una primera definición de serie de tiempo. Luego, presentaremos los distintos métodos y modelos más utilizados en el análisis de las series de tiempo, así como un modelo más avanzado basado en redes neuronales. Las ideas propuestas siguen muy de cerca la estructura presentada en Peña (2010) [15].

DEFINICIÓN 1.3. Sea $\{X_t\}_{t \in \mathbb{Z}}$ un proceso estocástico, denominaremos trayectoria o realización a la observación

$$\cdots, X_{-2}, X_{-1}, X_0, X_1, X_2, \cdots$$

Si consideramos una trayectoria parcial de un proceso estocástico de la forma

$$X_1, X_2, \cdots, X_T$$

esta recibe el nombre de serie de tiempo.

Las series de tiempo pueden ser divididas en dos tipos, series estacionarios y no estacionarias. Una serie de tiempo será estacionaria si cumple que:

- Las distribuciones marginales de todas las variables son idénticas.
- Las distribuciones de cualquier conjunto de variables solo dependen de la separación en tiempo (retardo).

Esta definición de proceso estacionario es muy estricta y de poca aplicabilidad en casos reales. Por este motivo se suele utilizar una forma de proceso estacionario menos estricta a la que se le denomina proceso estacionario en sentido débil. Una serie de tiempo es estacionaria en sentido débil si:

- $\mu_t = cte$
- $\sigma_t = cte$
- $E[(x_t - \mu)(x_{t+n} - \mu)] = E[(x_0 - \mu)(x_n - \mu)]$

3.1.1. Los modelos AR, MA y ARMA.

Cuando trabajamos con modelos estacionario los modelos más utilizados son los autorregresivos, los de media móvil, y los autorregresivos de media móvil. Estos modelos entran dentro de las herramientas clásicas del análisis de series de tiempo. Tienen la ventaja de que son aplicables en un abanico amplio de casos, aparte de ser necesarios para uno de los modelos más habituales de uso cuando no se cumple que el proceso es

estacionario, el modelo ARIMA.

El primer modelo a explicar será el autorregresivo (AR). De los modelos que se estudiarán este es el que tiene una mayor importancia, puesto que el trabajo se centra en cierta medida en una generalización de este al caso funcional.

DEFINICIÓN 1.4. Un proceso estacionario $\{X_t\}_t$ que admita la representación

$$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + a_t$$

Se conoce como proceso autorregresivo de orden p (AR(p)), siendo c, ϕ_1, \dots, ϕ_p constantes y a_t un proceso de ruido blanco¹.

El siguiente modelo es el de media móvil (MA).

DEFINICIÓN 1.5. Un proceso estacionario $\{X_t\}_t$ que admita la representación

$$X_t = c + a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \cdots + \theta_p a_{t-p}$$

Se conoce como proceso de media móvil de orden p (MA(p)), siendo $c, \theta_1, \dots, \theta_p$ constantes.

Cuando utilizamos ambas estructuras para describir un proceso estacionario de como resultado el proceso ARMA.

DEFINICIÓN 1.6. Un proceso estacionario $\{X_t\}_t$ que admita la representación

$$X_t = +\phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \cdots + \theta_q a_{t-q}$$

Se conoce como proceso ARMA(p, q), siendo $c, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$ constantes.

¹Un proceso de ruido blanco es un proceso estacionario que tiene media cero y es incorrelado.

3.1.2. Los modelos ARIMA.

Los modelos ARMA únicamente son aplicables en series de tiempo estacionarias. Lo cierto es que este tipo de series no abundan en la realidad, generalmente suelen presentar algún tipo de tendencia, patrones estacionales, heterocedasticidad, etc. La heterocedasticidad es un problema que puede ser resuelto con relativa facilidad transformando nuestros datos mediante las transformaciones de Box-Cox. La tendencia y estacionalidad requieren un tratamiento ligeramente distinto, buscando generalizar los procesos ARMA. De esta forma surgen los modelos ARIMA.

DEFINICIÓN 1.7. Un proceso ARIMA(p,d,q) es aquel proceso que requiere aplicar un número d de diferenciaciones² hasta alcanzar un modelo ARMA(p,q).

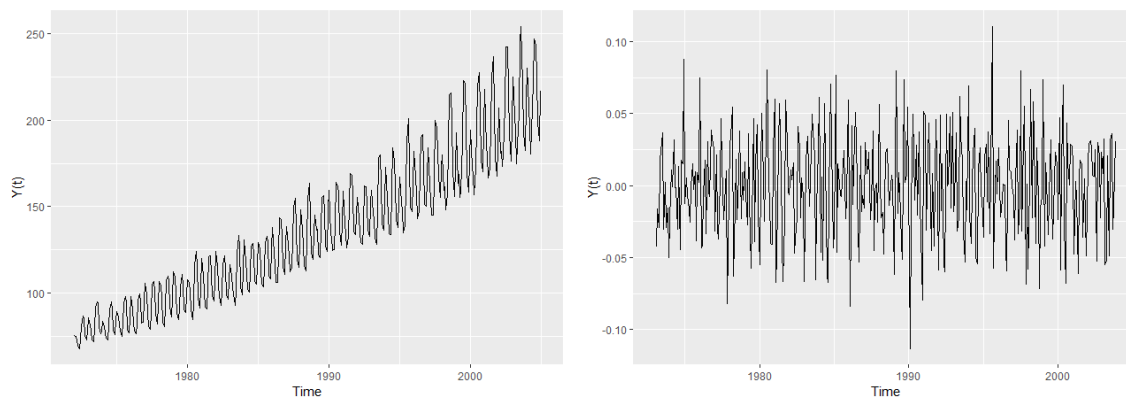


Figura 3.1: Serie de tiempo y serie de tiempo diferenciada

3.1.3. Red neuronal autorregresiva.

En los apartados anteriores se ha dado una breve descripción de los métodos más comunes y simples en el estudio de las series de tiempo. Existen una gran variedad de métodos más avanzados que permiten dar una mejor predicción de la serie de tiempo. Uno de ellos son las redes neuronales autorregresivas. Se dará ahora una breve descripción de su funcionamiento ya que será útil en futuras secciones cuando se extenderán estas ideas al caso funcional. Para una descripción más detallada Hyndman et al. (2018)[13]

Una red neuronal puede considerarse como una red que es organizada en capas. Las entradas de la primera capa son los predictores (o entradas), mientras que la predicción es la salida de la última capa. Pueden existir capas intermedias a las que se le da el

²Se entiende por diferenciación al método en el que calculamos un nuevo proceso a partir de la resta $Y_t = X_t - X_{t-1}$

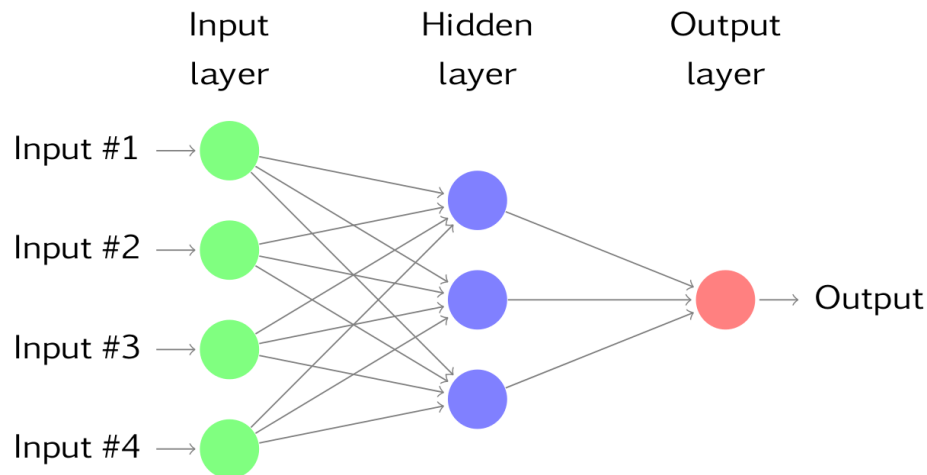


Figura 3.2: Red neuronal con una capa oculta.

nombre de capas ocultas. La idea de una red se basa ligeramente en el funcionamiento de las redes neuronales biológicas. En la imagen 3.2 se muestra una red neuronal con una capa oculta. Las salidas de los nodos en una capa son entradas a la siguiente capa. Las entradas a cada nodo son una combinación lineal ponderada de la forma

$$z_j = b_j + \sum_i w_{ij}x_i$$

Esta suma ponderada es evaluada sobre una función de activación, generalmente una función sigmoide como la función logística

$$s(z) = \frac{1}{1 + e^{-z}}$$

Los valores de esta función serán las nuevas entradas de la siguiente capa. Las ponderaciones de la red, usualmente denominadas pesos, son estimados mediante algoritmos de aprendizaje. El funcionamiento de estos algoritmos de aprendizaje no es más que la de minimizar una función error definida de forma acorde al problema.

Se puede aplicar esta metodología a las series de tiempo. Para entender un poco mejor el funcionamiento vamos a ejemplificar con un modelo simple. Imaginemos que tenemos un modelo AR(3). Este modelo es dependiente hasta el tercer retardo X_{t-3} . La idea aquí es utilizar los tres retardos como valores de entrada de la red y como salida obtener la predicción. A estos modelos se les da el nombre de red neuronal autorregre-

siva (NNAR).

3.2. Modelos de regresión paramétricas aplicados a las series de tiempo funcionales.

Durante todo el capítulo anterior hemos dado una descripción detallada de los modelos de regresión lineal, ahora nos centraremos en la aplicación al ámbito de las series de tiempo funcional. Para ello daremos una ligera modificación de los modelos anteriormente explicados. Comenzaremos con el caso más simple, la regresión con respuesta escalar.

3.2.1. Regresión con respuesta escalar para series de tiempo funcionales.

Antes de propiamente reformular el modelo de regresión con respuesta escalar al enfoque de las series de tiempo funcionales, vamos a justificar la utilización de esta metodología. Imagines que nuestro objetivo es predecir el comportamiento futuro de una variable que es escalar, un ejemplo que podríamos emplear sería el valor esperado de pacientes que acuden a un hospital condicionado a la temperatura de los días anteriores. Si tenemos un registro de las temperaturas a lo largo del día podríamos tratar de determinar el número de pacientes que acudirán al hospital padeciendo los síntomas del resfriado común. Esta situación entraría en el marco de las series de tiempo funcionales, ya que tratamos de estimar valores futuros a partir de valores pasados.

El modelo de regresión tiene una estructura similar al planteado en el capítulo anterior, pero con pequeñas diferencias.

$$y_{i+k} = \langle X_i, \beta \rangle + \epsilon_{i+k} = \int X_i(t)\beta(t)dt + \epsilon_{i+k}$$

Donde el índice k nos indica el evento k -ésimo. Es frecuente que el evento k -ésimo sea el siguiente, aunque se podrían dar situaciones en las que esto no sea cierto y fuera necesario trabajar con un retardo. Para contextualizar un poco esta idea vamos a retomar el ejemplo de los pacientes del resfriado y la temperatura. Si una persona está expuesta a una situación en la que la temperatura es baja, será más propensa a enfermarse por un resfriado. Sin embargo, los síntomas del resfriado no aparecen al día siguiente, sino que se requiere un cierto tiempo de incubación. Esto implica que la persona no acudirá al hospital inmediatamente, por lo que será necesario estudiar nuestra

serie de tiempo con un cierto retardo.

3.2.2. Regresión con respuesta funcional para series de tiempo funcionales.

Pasaremos ahora a detallar el funcionamiento de las series de tiempo mediante la regresión con respuesta funcional. En este caso podemos enfocar la regresión desde dos puntos de vista, el primero es aquel en el que nuestra covariable sea la misma que nuestra respuesta, como ocurre con los modelos ARMA de las series de tiempo clásicas. El segundo es en que la covariable y la variable respuesta son distintas, como nos ocurría con el caso de respuesta escalar. Para ambos casos el modelo se puede describir de la misma forma:

$$Y_{i+k}(t) = \int X_i(s)\beta(t, s)ds + \epsilon_{i+k}(t)$$

donde si $Y_{i+k} = X_{i+k}$ estamos en el primer caso, mientras que $Y_{i+k} \neq X_{i+k}$ es el segundo.

Para entender un poco mejor estos casos vamos a utilizar dos ejemplos simples. Imaginemos que estamos interesados en determinar el comportamiento de una serie de tiempo financiera, estudiándola desde un punto de vista funcional. Nuestro principal interés en este análisis es determinar cómo será el comportamiento en el futuro de esta serie a partir de cómo se comporta en el presente. Este ejemplo englobaría el primer caso, la covariable y la variable respuesta son la misma. Para el segundo caso podemos utilizar el consumo eléctrico de una ciudad. En principio podríamos utilizar el mismo análisis que en el caso anterior, no obstante, podría ser más útil en este caso utilizar otra variable distinta como podría ser la distribución de la temperatura.

3.3. Modelos no paramétricos.

Una de las metodologías más comúnmente utilizadas en las series de tiempo funcionales es el modelo de regresión no paramétrico. El modelo de regresión no paramétrico funcional puede extenderse a las series de tiempo funcionales de la siguiente manera.

$$X_{i+1}(t) = m(X_i) + \epsilon_{i+1}(t) \tag{3.1}$$

La ecuación 3.1 supone que los valores de la serie en el instante t en un intervalo dado es una función no paramétrica desconocida de la serie en todo el intervalo anterior

más algún término de error.

En nuestro estudio utilizaremos el estimador de Nadaraya-Watson de m basado en las ideas de Aneiros-Perez et al.(2011) [1]. Este modelo viene definido como

$$\hat{m}_h(X) = \sum_{i=1}^N \omega_h(X, X_i) X_{i+1}(t)$$

donde el parámetro h es la ventana de suavizado y ω

$$\omega_h(X, X_i) = \frac{K(d(X, X_i)/h)}{\sum_j K(d(X, X_j)/h)}$$

Donde K es una función tipo núcleo y d es la distancia entre X y X_i . Esta distancia d dependerá de la métrica o semimétrica elegida. Es aquí donde reside la principal diferencia entre la regresión no paramétrica clásica y la funcional.

El uso de un núcleo determinado no es en principio una cuestión que implique una gran relevancia en el modelo. Por otro lado, la ventana y la métrica utilizada si tendrán relevancia. En nuestro caso siempre trabajaremos con el espacio L^2 por lo que el punto clave del análisis del modelo no paramétrico será la ventana.

Existe una literatura extensa dedicada a la construcción de reglas de selección del ancho de ventana en los modelados no paramétrico. La mayoría de estas metodologías consisten en buscar los valores de la ventana que minimizan alguna medida de precisión de la estimación. En nuestro modelo, minimizaremos la suma residual de cuadrados:

$$MSE = \frac{1}{N} \sum_{i=1}^{N-1} \int (m(X_i(t)) - X_{i+1}(t))^2 dt$$

Una de las técnicas más utilizadas es la validación cruzada. Tiene como desventaja que su uso puede resultar computacionalmente costoso, mientras que su principal ventaja es la simplicidad a la hora de implementar en prácticamente cualquier ámbito. La ventana elegida será aquella que cumpla la minimización

$$\hat{h} = \min_h CV(h)$$

donde

$$CV(h) = \frac{1}{N} \sum_{i=1}^{N-1} \int (m^i(X_i(t)) - X_{i+1}(t))^2 dt$$

siendo $m^i(X)$ el estimador de Nadaraya-Watson estimado con todo el set de datos salvo la observación i -ésima.

3.4. Modelo de red neuronal en series de tiempo de datos funcionales.

Una de las herramientas más utilizadas a lo largo de los últimos años en el ámbito del aprendizaje estadístico han sido las redes neuronales. Al principio del capítulo ya se hizo una breve descripción de como estas metodologías se habían implementado en el estudio de las series de tiempo con las redes neuronales autorregresivas. Nuestro objetivo ahora será proponer un modelo de redes neuronales basándonos en las ideas de Portela (2017) [16] aunque ligeramente más simple. Antes de comenzar vamos a enunciar el teorema de representación de Kolmogorov-Arnold.

TEOREMA 3.1. Dado $n \in \mathbb{Z}$, toda función $f \in C([0, 1]^n)$ puede ser representado como

$$f(x_1, x_2, \dots, x_n) = \sum_{q=1}^{2n+1} g_q \left(\sum_{p=1}^n \phi_{pq}(x_p) \right)$$

donde $\phi_{pq} \in C[0, 1]$ son funciones crecientes independientes de f y $g_q \in C[0, 1]$ dependientes en f .

La idea cualitativa de todo nuestro análisis es simple, conseguir estimar el operador $\rho(t, s)$ que cumpla

$$X_{i+1}(t) = \int X_i(s) \rho(t, s) ds + \epsilon_{i+k}(t)$$

Nosotros propondremos ahora la estimación de este operador empleando el teorema 3.1. Definiremos el estimador de nuestro operador como la suma de la siguiente forma

$$\rho(t, s) = \sum_{j=1}^c \omega_{0j} f(\omega_{1j} + \omega_{2j}t + \omega_{3j}s)$$

donde los parámetros ω_{ij} serán pesos a estimar y la función f es una función sigmoide como podría ser la función logística, la tangente hiperbólica o cualquier otra. En nuestro análisis usaremos la función logística

$$f(\omega_{1j} + \omega_{2j}t + \omega_{3j}s) = \frac{1}{1 + e^{-(\omega_{1j} + \omega_{2j}t + \omega_{3j}s)}}$$

En la imagen 3.3 se muestra como la suma de estas sigmoides bidimensionales estiman la función $\rho(s, t)$.

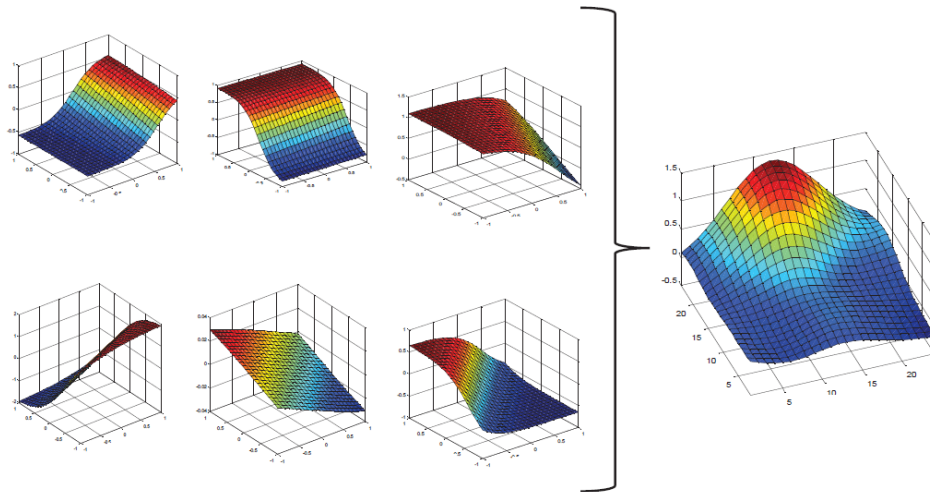


Figura 3.3: Suma de funciones logísticas bidimensionales para estimar ρ

Este modelo se puede ver como una red neuronal que toma como entradas las variables t y s continuas, una capa oculta cuyo número de nodos vendrá dada por el valor c elegido en el desarrollo y un único nodo de salida que será la suma de los valores obtenidos en los nodos. En la imagen 3.4 se muestra el diagrama de nuestro modelo en el caso de que nuestra $c = 5$.

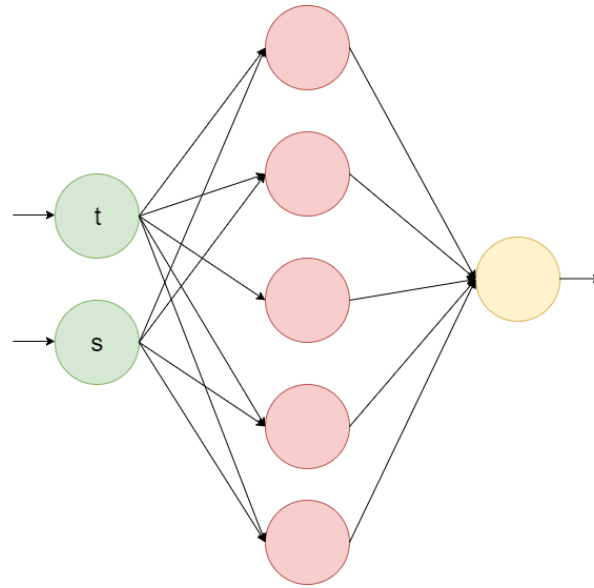


Figura 3.4: Red neuronal para estimar ρ con cinco nodos

3.4.1. Estimación de los parámetros.

En esta sección daremos una descripción de los métodos empleados para estimar los parámetros de modelo. Utilizaremos una de las técnicas más utilizadas y simples, el descenso del gradiente. Como primer punto de partida debemos definir una función de error a minimizar. Nosotros utilizaremos la suma del cuadrado de las distancias entre la variable real y nuestra predicción.

$$E = \sum_{i=1}^{N-1} \int (X_i(t) - \widehat{X}_i)^2 dt$$

Nuestra función E tomara un mínimo para un parámetro ω_{ij} dado en aquellos valores en los que su derivada sea cero. Es conveniente escribir la derivada de E en la siguiente forma

$$\frac{\partial E}{\partial \omega_{ij}} = \frac{\partial E}{\partial \widehat{X}_i} \frac{\partial \widehat{X}_i}{\partial \omega_{ij}}$$

donde hemos utilizado la regla de la cadena. Nos queda por tanto que los valores que minimizan nuestro error son aquellos que cumplen

$$\frac{\partial E}{\partial \omega_{ij}} = \sum_{i=1}^{N-1} \int -2 \left(X_i(t) - \hat{X}_i \right) \frac{\partial \hat{X}_i}{\partial \omega_{ij}} dt = 0$$

Como se mencionó anteriormente trabajaremos con la función logística bidimensional, la expresión pasa a ser

$$\begin{aligned} \frac{\partial E}{\partial \omega_{ij}} &= \sum_{i=1}^{N-1} \int -2 \left(X_i(t) - \hat{X}_i \right) \frac{\partial}{\partial \omega_{ij}} \left(\int \sum_{j=1}^c \tilde{\rho}^c(t, s) X_{i-1}(s) ds \right) dt = \\ &\sum_{i=1}^{N-1} \int -2 \left(X_i(t) - \hat{X}_i \right) \frac{\partial}{\partial \omega_{ij}} \left(\int \sum_{j=1}^c \omega_{0j} \frac{1}{1 + e^{-(\omega_{1j} + \omega_{2j}t + \omega_{3j}s)}} X_{i-1}(s) ds \right) dt \end{aligned}$$

donde las derivadas para los distintos parámetros son

$$\begin{aligned} \frac{\partial \rho(t, s)}{\partial \omega_{0j}} &= \frac{1}{1 + e^{-(\omega_{1j} + \omega_{2j}t + \omega_{3j}s)}} \\ \frac{\partial \rho(t, s)}{\partial \omega_{1j}} &= \omega_{0j} \frac{e^{-(\omega_{1j} + \omega_{2j}t + \omega_{3j}s)}}{(1 + e^{-(\omega_{1j} + \omega_{2j}t + \omega_{3j}s)})^2} \\ \frac{\partial \rho(t, s)}{\partial \omega_{2j}} &= \omega_{0j} \frac{e^{-(\omega_{1j} + \omega_{2j}t + \omega_{3j}s)}}{(1 + e^{-(\omega_{1j} + \omega_{2j}t + \omega_{3j}s)})^2} t \\ \frac{\partial \rho(t, s)}{\partial \omega_{3j}} &= \omega_{0j} \frac{e^{-(\omega_{1j} + \omega_{2j}t + \omega_{3j}s)}}{(1 + e^{-(\omega_{1j} + \omega_{2j}t + \omega_{3j}s)})^2} s \end{aligned}$$

Una vez que hemos definido la derivada de nuestra función error trataremos de buscar su mínimo utilizando la técnica del descenso del gradiente. Se trata de un método iterativo donde el siguiente valor de nuestro parámetro es

$$\omega_{k+1} = \omega_k - \gamma \nabla E(\omega_k)$$

Donde el parámetro γ toma siempre valores mayores que cero. En el ámbito del aprendizaje estadístico se le da el nombre de *learning rate* pues determina la velocidad de convergencia del método. Si elegimos valores muy pequeños el algoritmo requerirá un gran número de pasos para poder determinar el valor del mínimo, mientras que valores

muy altos causaran la no convergencia del método. Es por este motivo que determinar un valor adecuado es considerablemente importante. En la imagen 3.5 se muestra de forma esquemática el funcionamiento del descenso del gradiente. En este caso se representa el caso más simple, en el tenemos como objetivo encontrar el mínimo de una función con un solo mínimo. En nuestro modelo el caso se complica ya que tenemos un número alto de mínimos locales. El descenso del gradiente no es capaz de determinar si se trata de un mínimo local o global, simplemente alcanzara aquel mínimo que se encuentre más cerca del primer valor inicial dado. Por este motivo es común que se utilicen distintos puntos de partida a la hora de determinar el mínimo.

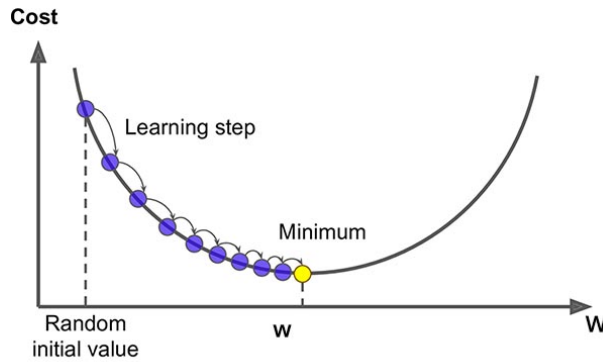


Figura 3.5: Esquema del descenso del Gradiente

Vamos ahora es reescribir el problema en forma matricial, definiremos las matrices ω^k y J como

$$\omega^k = \begin{bmatrix} \omega_{00}^k & \omega_{10}^k & \omega_{20}^k & \omega_{30}^k \\ \omega_{01}^k & \omega_{11}^k & \omega_{21}^k & \omega_{31}^k \\ \vdots & \vdots & \vdots & \vdots \\ \omega_{0c}^k & \omega_{1c}^k & \omega_{2c}^k & \omega_{3c}^k \end{bmatrix}, J = \begin{bmatrix} \frac{\partial \rho(t,s)}{\partial \omega_{00}} & \frac{\partial \rho(t,s)}{\partial \omega_{10}} & \frac{\partial \rho(t,s)}{\partial \omega_{20}} & \frac{\partial \rho(t,s)}{\partial \omega_{30}} \\ \frac{\partial \rho(t,s)}{\partial \omega_{01}} & \frac{\partial \rho(t,s)}{\partial \omega_{11}} & \frac{\partial \rho(t,s)}{\partial \omega_{21}} & \frac{\partial \rho(t,s)}{\partial \omega_{31}} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial \rho(t,s)}{\partial \omega_{0c}} & \frac{\partial \rho(t,s)}{\partial \omega_{1c}} & \frac{\partial \rho(t,s)}{\partial \omega_{2c}} & \frac{\partial \rho(t,s)}{\partial \omega_{3c}} \end{bmatrix}_{\omega=\omega^k}$$

Siendo ω^k la matriz de parametros en la k -ésima iteración y J la matriz jacobiana evaluada en ω^k . De esta forma podemos reescribir nuestro algoritmo iterativo como

$$\omega_{k+1} = \omega_k - \gamma J$$

Uno de los problemas más comunes en el ámbito del aprendizaje estadístico es el denominado *overfitting*. Cuando hablamos de *overfitting* nos referimos al sobreajuste que ciertos modelos pueden generar debido a que tienden a ajustarse de forma excesiva al set de datos original perdiendo así flexibilidad. Las redes neuronales son métodos que pueden sufrir este problema con relativa facilidad. Para evitarlo generalmente se divide el set de datos original en dos partes. La primera parte será el set de datos que se utiliza para entrenar el modelo, mientras que con la segunda se realiza el test de evaluación. Como los datos del test no han sido utilizados en el entrenamiento no sufrirán problemas de *overfitting*. De esta forma determinaremos la calidad del ajuste real.

Capítulo 4

Modelo autorregresivo en espacios de Hilbert.

En los capítulos anteriores nos hemos centrado en los modelos de regresión funcionales tanto paramétricos como no paramétricos. Existe otra metodología distinta propuesto por Bosq (2000)[5], donde se hace una extensión de los modelos autorregresivos de series de tiempo clásicos al caso funcional. Esta metodología recibe el nombre de modelo autorregresivo en espacios de Hilbert. Comenzaremos analizando el caso más simple, el modelo autorregresivo en espacios de Hilbert de orden 1 ($ARH(1)$).

4.1. Modelo autorregresivo en espacios de Hilbert de orden 1.

Como punto de partida comenzaremos dando la definición del modelos autorregresivos en espacios de Hilbert de orden 1 y de la extensión del ruido blanco al caso funcional.

DEFINICIÓN 4.1. Sea $X = (X_n, n \in \mathbb{Z})$ una secuencia de variables aleatorias funcionales definidas en un espacio de Hilbert H , es un proceso $ARH(1)$ si es estacionario y cumple que

$$X_n = \rho(X_{n-1}) + \epsilon_n, \quad n \in \mathbb{Z} \tag{4.1}$$

donde $\mathbb{E}[X] = 0$ y ϵ es un H-ruido blanco.

DEFINICIÓN 4.2. Una secuencia $\epsilon = (\epsilon_n, n \in \mathbb{Z})$ de variables aleatorias funcionales definidas en un espacio de Hilbert H , es denominado H-ruido blanco si

1. Tiene media cero $E[\epsilon_n] = 0$
2. Tiene varianza finita $0 < \mathbb{E}\|\epsilon_n\|^2 = \sigma^2 < \infty$
3. El operador de la covarianza no depende de n $C_\epsilon = C_{\epsilon_n}$
4. ϵ_n es ortogonal a ϵ_m

La estructura de este modelo es idéntica a la de los modelos de regresión de los capítulos anteriores, con la diferencia de que en este caso estamos imponiendo condiciones ligeramente más restrictivas que en los casos anteriores. Estas condiciones son necesarias para poder describir la serie de tiempo de desde este punto de vista.

Bajo este análisis entra en juego una cuestión de gran relevancia que en los anteriores análisis no nos habíamos planteado. Determinar en qué condiciones podemos implementar los modelos de autorregresivos en espacios de Hilbert o lo que es lo mismo, las condiciones de existencia.

4.1.1. Existencia.

Para contextualizar el problema de la existencia vamos a estudiar brevemente que ocurre con el modelo autorregresivo de orden uno clásico (AR(1)). Recordando la definición 3.4 dada en el primer capítulo, sea X_t un proceso estacionario clásico, diremos que admite AR(1) si ha sido generado por:

$$X_t = c + \phi X_{t-1} + a_t$$

donde a_t es un proceso de ruido blanco, c es una constante y ϕ es otra constante que toma valores entre $(-1, 1)$.

Si asumimos que la serie X_t comienza en el infinito pasado podemos representar el proceso como:

$$X_t = \sum_{i=0}^{\infty} \phi^i a_{t-i}$$

Que se trata de una serie geométrica que será convergente únicamente si $|\phi| < 1$. A los procesos AR(1) que admiten esta representación se les denomina causales.

4.1. MODELO AUTORREGRESIVO EN ESPACIOS DE HILBERT DE ORDEN 1.43

Para el caso funcional la situación es similar, aunque ligeramente más compleja ya que trabajamos con operadores y no con constantes. Nuestro objetivo es establecer una condición análoga a $|\phi| < 1$ para el operador ρ .

LEMA 4.1. Para todo $\rho \in \mathcal{L}$ las siguientes dos condiciones son equivalentes.

c1) Existe un entero $j_0 \geq 1$ tal que $\|\rho^{j_0}\|_{\mathcal{L}} < 1$

c2) Existe un $a > 0$ y $0 < b < 1$ tal que $\|\rho^j\|_{\mathcal{L}} < ab^j, j \geq 0$

La demostración completa del lema se adjunta en el apéndice.

TEOREMA 4.1. Si la condición c1) del lema es cierta, entonces la ecuación 4.1 tiene una única solución estacionaria que viene dada por

$$X_n = \sum_{i=0}^{\infty} \rho^i(\epsilon_{n-i}), \quad n \in \mathbb{Z}$$

La serie converge de forma casi segura en L^2 , es decir

$$E\|X_n - \sum_{i=0}^l \rho^i(\epsilon_{n-i})\| \rightarrow 0, \quad l \rightarrow \infty$$

La demostración se añade en el apéndice.

Un ejemplo relevante de operadores que cumplen esta condición es el de los operadores integrales de Hilbert-Schmidt, que venían definidos como

$$\Psi(t) = \int \psi(t, s)x(s)ds, \quad x \in L^2$$

y satisface

$$\int \int \psi^2(t, s)dt ds < 1$$

que es precisamente $\|\Psi\|_S^2$. Sin embargo, en la sección de análisis de datos funcionales del primer capítulo, se mencionó que $\|\Psi\|_{\mathcal{L}} \leq \|\Psi\|_S$, lo que implica la condición c1 del lemma para $j_0 = 1$.

4.1.2. Estimación.

Una vez que hemos establecido las condiciones en las que podemos tratar de estudiar la serie de tiempo desde el punto de vista de los modelos autorregresivos, el siguiente objetivo es estimar el operador $\rho(x)$. Sobre él se imponen dos condiciones, que sea simétrico y definido positivo. La idea de estas dos condiciones es simple, poder utilizar la descomposición espectral.

$$X_{n+1} = \rho(X_n) + \epsilon_{n+1}$$

Para estimar este operador primero aplicaremos

$$E[\langle X_{n+1}, x \rangle X_n] = E[\langle \rho(X_n), x \rangle X_n] + E[\langle \epsilon_{n+1}, x \rangle X_n]$$

Pero como ϵ es incorrelado en H y ρ es simétrico tenemos

$$E[\langle X_{n+1}, x \rangle X_n] = E[\langle \rho(X_n), x \rangle X_n] = E[\langle X_n, \rho(x) \rangle X_n] = C(\rho^T(x))$$

Donde C no es más que el operador de la covarianza. Definimos el operador D como

$$D(x) = E[\langle X_n, x \rangle X_{n+1}]$$

siendo $D^T = E[\langle X_{n+1}, x \rangle X_n]$. Este operador D recibe el nombre de covarianza cruzada. Tenemos por tanto que $D^T = C(\rho^T)$, es decir

$$D(x) = \rho(C(x))$$

4.1. MODELO AUTORREGRESIVO EN ESPACIOS DE HILBERT DE ORDEN 1.45

Con lo que nuestro objetivo se reduce únicamente a encontrar la inversa del operador de covarianzas $C(x)$. Como se mostró en el primer capítulo un estimador de $C(x)$ es ¹

$$\widehat{C}(x) = \frac{1}{n} \sum_{i=1}^n \langle X_i, \cdot \rangle X_i$$

Sin embargo, la dimensión de nuestro espacio es infinita causando que no necesariamente se cumpla $\rho = DC^{-1}$ en todo el espacio. Por este motivo definiremos el operador de covarianza utilizando la descomposición espectral, limitando el desarrollo hasta un número K . De esta forma conseguimos obtener una aproximación de la inversa.

$$\widehat{C}_n \simeq \sum_i^K \widehat{\lambda}_{jn} \langle \nu_i, \cdot \rangle \nu_j$$

donde

$$\widehat{\lambda}_{in} = \frac{1}{n} \sum_j^n \langle X_j, \nu_j \rangle^2, \quad j \geq 1; n \geq 1$$

con

$$\sum_{i=1}^K \widehat{\lambda}_{in} < \infty$$

Este estimador si es invertible sobre todo H con inversa

$$\widehat{C}_n^{-1} \simeq \sum_i^K \widehat{\lambda}_{jn}^{-1} \langle \nu_i, \cdot \rangle \nu_j$$

Por otro lado el estimador del operador D es

¹En este caso estamos asumiendo que $E[X] = 0$

$$\widehat{D}(x) = \frac{1}{n-1} \sum_{i=1}^n \langle X_i, \cdot \rangle X_{i+1}$$

Una vez determinado la inversa del estimador de C y el estimador D ya podemos determinar un estimador del operador $\widehat{\rho}$

$$\widehat{\rho}_n(x) = \frac{1}{n-1} \sum_{l=1}^{k_n} \sum_{i=1}^{n-1} \sum_{j=1}^{k_n} \widehat{\lambda}_{jn}^{-1} \langle x, \nu_j \rangle \langle X_i, \nu_j \rangle \langle X_{i+1}, \nu_l \rangle \nu_l$$

Estimación de las autofunciones.

Un pilar clave en la metodología antes expuesta, es el uso y la determinación de los autofunciones para estimar nuestro operador. Clásicamente la metodología habitual es utilizar las componentes principales como autofunciones. Tienen la ventaja de que con un pequeño número de elementos de la base ya podemos abarcar una gran variabilidad. Sin embargo, esta metodología no se enfoca en determinar aquellas autofunciones que sean más útiles a la hora de predecir. Una alternativa interesante podría ser utilizar las componentes PLS en lugar de las componentes principales, ya que estas si se enfocan en determinar aquellas autofunciones que sean más útiles en la predicción. En este trabajo únicamente nos centraremos en el uso de las componentes principales.

4.1.3. Ejemplo ilustrativo.

En este apartado vamos a tratar de comprobar como de buena es esta estimación del operador ρ utilizando componentes principales. Para ellos utilizaremos un ejemplo ilustrativo.

Antes de comenzar vamos a reescribiremos ligeramente $\widehat{\rho}$ como un operador integral. De esta forma generaremos una serie de tiempo utilizando un operador ρ integral con núcleo β y lo compararemos con el núcleo estimado por el procedimiento descrito anteriormente. Definiremos $\widehat{\beta}(t, s)$ como

$$\widehat{\beta}(t, s) = \frac{1}{n-1} \sum_{l=1}^{k_n} \sum_{i=1}^{n-1} \sum_{j=1}^{k_n} \widehat{\lambda}_{jn}^{-1} \langle X_i, \nu_j \rangle \langle X_{i+1}, \nu_l \rangle \nu_l \nu_j$$

4.1. MODELO AUTORREGRESIVO EN ESPACIOS DE HILBERT DE ORDEN 1.47

que cumple

$$\widehat{\rho}(x)(t) = \int \widehat{\beta}(t, s)x(s)ds$$

Como se mencionó anteriormente una de las grandes ventajas de las componentes principales es que el uso de un pequeño número de ellas nos permiten dar una buena descripción del modelo. En este ejemplo utilizaremos únicamente las tres primeras componentes principales. Al igual que en el estudio teórico planteado para los modelos de regresión funcional, trabajaremos con innovaciones que vendrán dadas como una combinación lineal de funciones trigonométricas de la forma $\epsilon(t) = 0,5 * N_t(0, 1) * \text{sen}(\pi t) + 0,5 * N_t(0, 1) * \text{sen}(\pi t)$. En la imagen 4.1 se muestran en azul el valor teórico de β mientras que las otras tres gráficas son las estimaciones utilizando 1, 2 y 3 componentes principales.

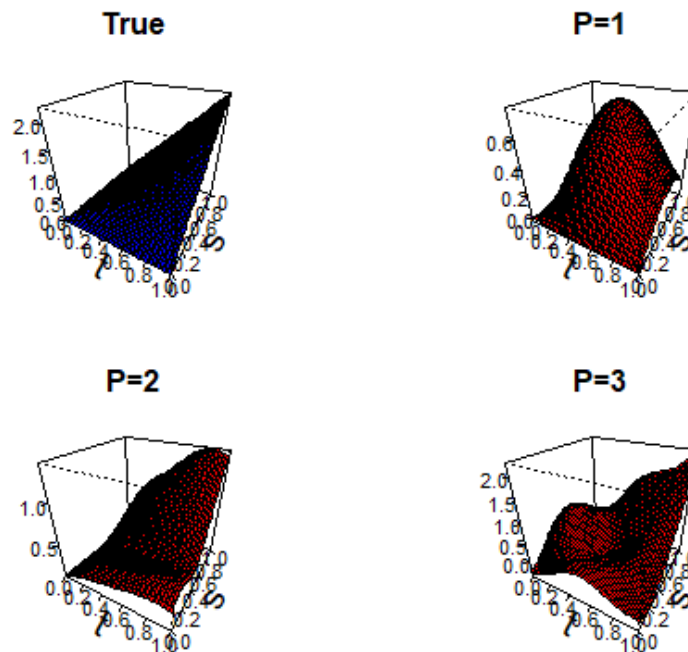


Figura 4.1: Superficie del núcleo teórica y estimada para las tres primeras componentes principales

A medida que aumentamos el número de componentes vemos como la escala se aproxima más a la original, pero en contraposición tiende a ser una superficie más rugosa. En términos generales la aproximación es bastante buena.

4.2. Modelo ARH(p).

El siguiente paso en el estudio de los modelos autorregresivos en espacios de Hilbert es dar una estructura del caso general, el modelo autorregresivo de orden p (ARH(p))

DEFINICIÓN 4.3. Sea $X = (X_n, n \in \mathbb{Z})$ una secuencia de variables aleatorias funcionales definidas en un espacio de Hilbert H , es un proceso $ARH(p)$ si es estacionario y cumple que

$$X_n = \rho_1(X_{n-1}) + \cdots + \rho_p(X_{n-p})\epsilon_n, \quad n \in \mathbb{Z}$$

donde $\mathbb{E}[X] = 0$ y ϵ es un H-ruido blanco.

La metodología utilizada para el caso de orden 1 no es aplicable en este caso puesto que en lugar de tener una ecuación, tenemos un sistema de ecuaciones donde nuestras incógnitas son operadores. Para resolver este problema trataremos de reducirlo al caso de orden 1. Comenzaremos dando la siguiente definición.

DEFINICIÓN 4.4. Denominaremos H^p al producto cartesiano de p copias de H , donde H^p viene dotado por el producto escalar

$$\langle (x_1, \cdots, x_p), (y_1, \cdots, y_p) \rangle_p = \sum_{i=1}^p \langle x_i, y_i \rangle$$

con $x_1, \cdots, x_p, y_1, \cdots, y_p \in H$

Sea $Y_n = (X_n, \cdots, X_{n-p+1})$ donde $E[X_i] = 0$, $\epsilon' = (\epsilon, 0, \cdots, 0)$ con tamaño p definimos el siguiente operador en H^p como

$$\rho' = \begin{bmatrix} \rho_1 & \rho_2 & \cdots & \rho_p \\ I & 0 & \cdots & 0 \\ 0 & I & \cdots & 0 \\ 0 & \cdots & I & 0 \end{bmatrix}$$

donde I es el operador identidad.

LEMA 4.2 Si X sigue un $ARH(p)$ asociado con $(\epsilon, \rho_1, \dots, \rho_p)$, entonces Y es un $ARH^p(1)$ asociado con (ϵ', ρ') .

Llegados a este punto únicamente tendremos que resolver el problema equivalente al del modelo $ARH(1)$ en el espacio H^p . Nuestra estimación de los operadores $\widehat{D}_y(x)$ y $\widehat{C}_{yn}(x)$

$$\widehat{C}_{yn}^{-1} \simeq \sum_i^{\infty} \widehat{\lambda}_{jn}^{-1} \langle \nu_i, \cdot \rangle_p \nu_j$$

con

$$\widehat{\lambda}_{in} = \frac{1}{n} \sum_j^n \langle Y_j, \nu_j \rangle_p^2, \quad j \geq 1; n \geq 1$$

y

$$\widehat{D}_y(x) = \frac{1}{n} \sum_{i=1}^n \langle Y_i, \cdot \rangle_p Y_{i+1}$$

Este análisis es idéntico al planteado por Bosq (2000)[5]. Una visión alternativa del modelo $ARH(p)$ se da en Kokoska et al. (2012)[14] donde también proponen una metodología para la determinación y contraste sobre el orden del modelo.

4.3. Modelo ARHX(p).

A lo largo de las secciones anteriores hemos estudiado como es el comportamiento de una serie de tiempo tratada como un objeto funcional. La cuestión que se nos plantea ahora es si existe alguna otra serie de tiempo que tenga influencia sobre la que estamos estudiando. Como ejemplo podríamos pensar en el precio de las acciones de dos bancos o el consumo de electricidad y la temperatura. La relevancia de estas relaciones entre series es muy grande, pues implica el tener más información a la hora de estudiar nuestro modelo. Es por este motivo que surge la idea de añadir nuevas variables al modelo, variables exógenas.

Nos enfocaremos en el estudio de los modelos autorregresivos en espacios de Hilbert con variables exógenas, estudiándolos como una "ampliación" de los ARH(p). Decimos como una ampliación puesto que la representación utilizada será equivalente a la de los modelos ARH(p).

Sea $\rho_1, \dots, \rho_p, a_1, \dots, a_l$ operadores lineales definidos sobre H, diremos que sigue un modelo autorregresivo en espacios de Hilbert con variables exógenas de orden p (ARHX(p)) si cumple que:

$$X_{n+1} = \rho_1(X_{n-1}) + \dots + \rho_p(X_{n-p}) + a_1(Z_{n,l}) + \dots + a_q(Z_{n,q}) + \epsilon_n$$

donde ϵ es ruido blanco hilbertiano, $Z_{n,1}, \dots, Z_{n,q}$ son variables que siguen un modelo autorregresivo de orden 1, es decir $Z_{n,i} = u(Z_{n-1,i}) + \epsilon_{n,i}$. Al igual que hicimos para el caso ARH(p) definimos un espacio de Hilbert separable H^q que tiene definido el producto escalar $\langle (x_1, \dots, x_q), (y_1, \dots, y_q) \rangle = \sum \langle x_i, y_i \rangle$. Si tenemos $Y_n = (X_n; \dots, X_{n-p+1}, Z_{n+1,1}, \dots, Z_{n+1,q})$ y $\epsilon = (\epsilon_n, 0, \dots, 0, \epsilon_{n,1}, \dots, \epsilon_{n,q})$ entonces podemos definir un operador ρ' como hicimos en el caso ARH(p) de la siguiente forma:

$$\rho' = \begin{bmatrix} \rho_1 & \cdots & \cdots & \rho_p & a_1 & \cdots & \cdots & a_p \\ I & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & I & 0 & \cdots & \cdots & \cdots & \cdots & \vdots \\ 0 & \cdots & I & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \cdots & \cdots & 0 & u_1 & 0 & \cdots & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & u_2 & \cdots & \vdots \\ \vdots & \cdots & \cdots & \cdots & 0 & \cdots & \ddots & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & \cdots & \cdots & u_q \end{bmatrix}$$

Cuando el proceso X_n siga un modelos ARHX(p) entonces Y_n sigue un modelo ARH(1) en H^q .

$$T_n = \rho'(T_{n-1}) + \epsilon'_n$$

Una descripción más detallada de este modelo se da en Damon et al.(2005) [6] donde se realiza un análisis utilizando muestras simuladas. Cabe destacar el trabajo Damon et al.(2002) [7] de los mismos autores que el anterior, donde utilizan los modelos autorregresivos con variables exógenas para predecir niveles de ozono. En este mismo trabajo se comparan los resultados de este modelo con los del modelo ARH(1) y un modelo de regresión no paramétrico funcional siendo el modelo ARHX el que mejores resultados obtiene.

Capítulo 5

Aplicación a datos reales.

En este capítulo nos centraremos aplicar las distintas metodologías recogidas a lo largo del trabajo a un caso con datos reales. En concreto utilizaremos las técnicas paramétricas, no paramétricas y los modelos autorregresivos en espacios de Hilbert de orden 1. Como set de datos a analizar utilizaremos el precio del intercambio entre el dólar y euro. Nuestro objetivo es comparar los resultados de las distintas metodologías, con el fin de determinar cuál es el más indicado para este caso en particular.

Surge por tanto una duda, como determinar qué modelo es el más adecuado. Para solucionar este problema utilizaremos distintas medidas de error.

5.1. Medidas de error.

En esta sección se dará una breve descripción de las distintas medidas de precisión, que se pueden calcular para los residuos de un modelo de predicción en el marco de las series de tiempo de datos funcionales. En general, no existe un criterio estándar para la medida de errores y generalmente depende del tipo de datos y la aplicación. En este trabajo utilizaremos tres de las medidas más habituales en este tipo de análisis:

- El error absoluto medio funcional (FMAE) definido como

$$FMAE = \frac{1}{N} \sum_{n=1}^T \|X_n - \hat{X}_n\|_{L^1} = \frac{1}{N} \sum_{t=1}^T \int |X_n(t) - \hat{X}_n(t)| dt$$

- El error cuadrático medio funcional (FMSE) definido como

$$FMSE = \frac{1}{N} \sum_{n=1}^T \|X_n - \hat{X}_n\|_{L^2}^2 = \frac{1}{N} \sum_{t=1}^T \int (X_n(t) - \hat{X}_n(t))^2 dt$$

- Si aplicamos la raíz a FMSE tenemos FRMSE

$$FRMSE = \sqrt{\frac{1}{N} \sum_{n=1}^T \|X_n - \hat{X}_n\|_{L^2}^2} = \sqrt{\frac{1}{N} \sum_{t=1}^T \int (X_n(t) - \hat{X}_n(t))^2 dt}$$

Estas mediciones de error pueden proporcionar algunos resultados que son difíciles de interpretar porque la integración se realiza en el rango de dominio de las funciones $t \in [a, b]$. Una forma de normalizar los resultados es dividiendo cada medida por $a - b$ como se muestra en Aneiros-Perez et al. (2013)[2].

5.2. Aplicación de los modelos con respuesta funcional.

Nuestro set de datos consiste en el precio del intercambio entre el dólar y el euro desde la formación de este último (Imagen 5.1). Una explicación más detallada sobre el set de datos se dará en apéndice C.

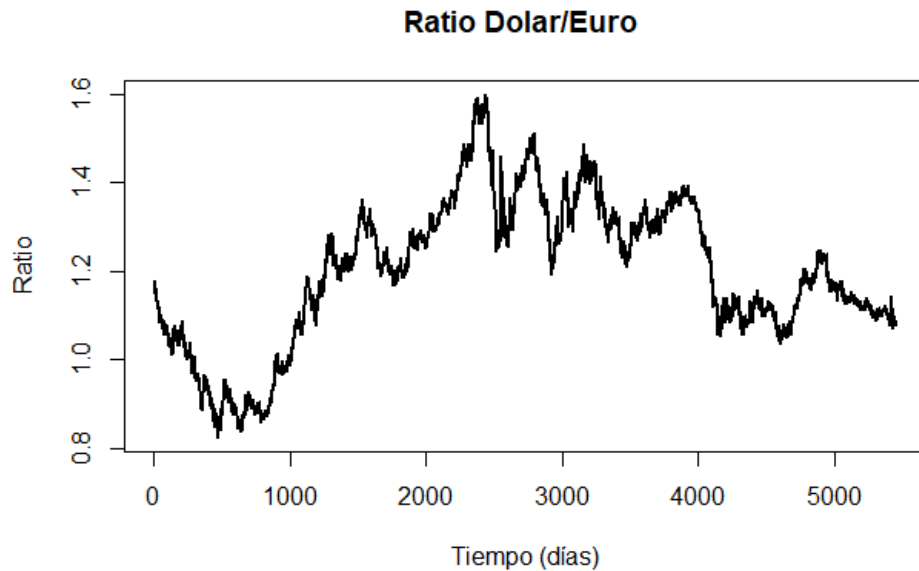


Figura 5.1: Relación entre el precio del dólar y el euro.

El objetivo del análisis es estudiar el comportamiento de la serie de tiempo utilizando las herramientas estudiadas en los capítulos anteriores. Trataremos cada dato funcional como la función que se estima a partir de los datos de 30 días (un mes). Utilizaremos

una colección grande de datos funcionales (en torno a los 150 primeros meses) para ajustar el modelo, y los siguientes meses restantes como test. Nuestra respuesta en este análisis será funcional y a diferencia de lo que se suele hacer en los modelos de series de tiempo clásicos, nuestras predicciones serán únicamente a un paso ¹. El imponer esta condición viene del hecho de que al trabajar con respuesta funcional no es en principio posible generar un intervalo de confianza para la predicción. Esto produce que las predicciones a largo plazo carezcan de sentido. En el caso de las series de tiempo clásicas no tenemos este problema, puesto que siempre podríamos recurrir a técnicas bootstrap para determinar los intervalos de confianza. Cabe destacar que este tipo de series de tiempo pueden ser tratadas desde otros enfoques completamente distintos. Entre ellos podemos destacar las ecuaciones diferenciales estocásticas, ampliamente utilizadas en los estudios financieros.

5.2.1. Aplicación del modelo de regresión paramétrico.

El primero modelo que estudiaremos será precisamente el primero detallado en el trabajo, el modelo de regresión funcional con respuesta funcional paramétrico. Utilizaremos la versión penalizada del modelo propuesto por Ramsay que se detalló en el capítulo dos.

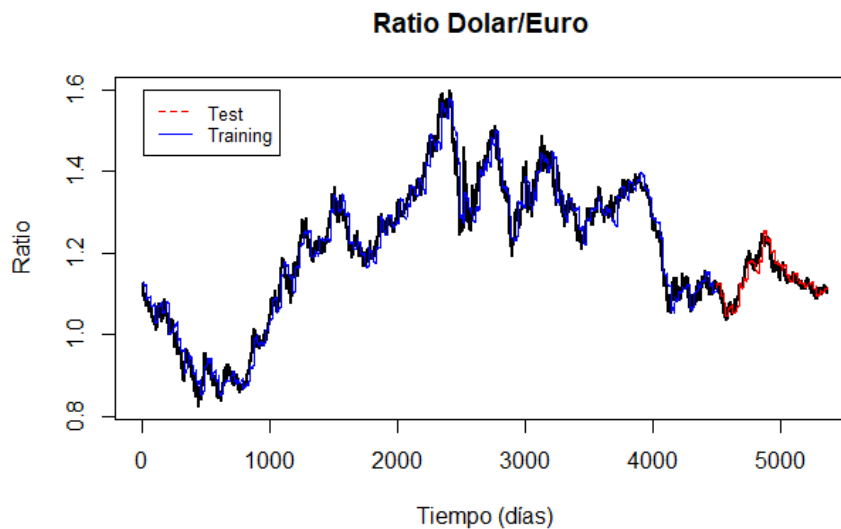


Figura 5.2: Estimación del modelo de regresión paramétrico.

¹Con predicciones a un paso nos referimos a que para realizar la predicción de un mes utilizamos el valor real del mes anterior. Esto quiere decir que si hemos hecho ya una predicción del mes de febrero y queremos realizar una para marzo, utilizaremos el valor real del mes de febrero y no la predicción.

En la imagen 5.2 se muestra el ajuste del modelo de regresión paramétrico. En azul se representa la estimación del set de datos utilizado para el ajuste, mientras que en rojo el conjunto de set de datos utilizados como test. El ajuste es considerablemente bueno, siendo ligeramente menos sensible a los cambios abruptos de tendencia.

5.2.2. Aplicación del modelo de regresión no paramétrico.

Nuestro siguiente modelo de estudio es el modelo de regresión funcional con respuesta funcional no paramétrico. La principal ventaja que tiene este modelo frente al anterior es que es más general, puesto que no asume ninguna condición de linealidad. Para nuestro estudio nos basaremos en las ideas de Ferraty, desarrolladas en el capítulo dos y tres.

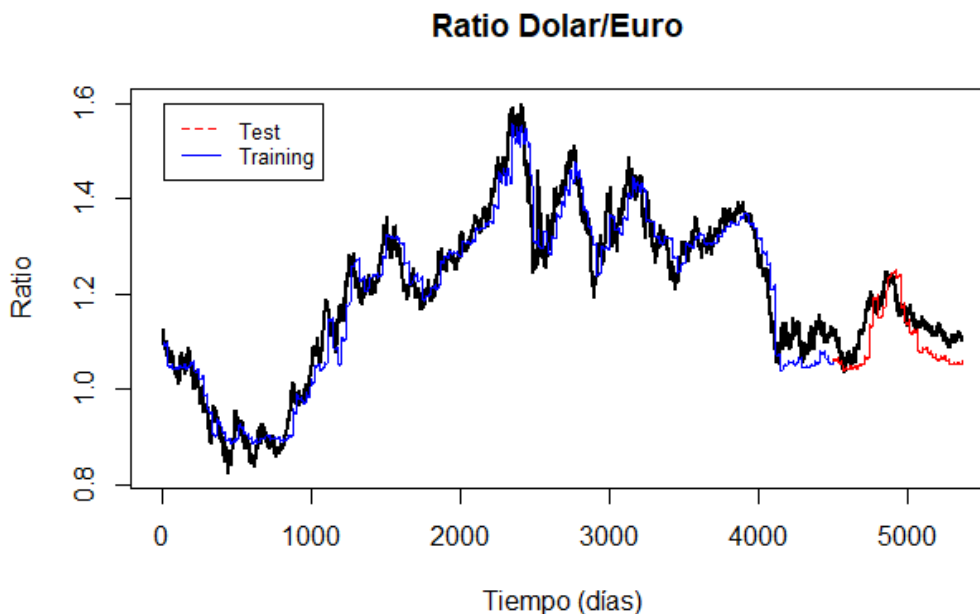


Figura 5.3: Estimación del modelo de regresión no paramétrico.

Como se puede ver en la imagen 5.3 el modelo de regresión no paramétrica tiene un resultado significativamente peor que el paramétrico. Si bien es cierto que acierta en la tendencia de la serie el resultado no es tan bueno como cabría esperar, más teniendo en cuenta que se trata de un modelo más flexible. Este resultado quizás pueda deberse a una selección de la ventana no tan exhaustiva como es necesario. Otra posibilidad es el uso de la medida de error FMSE para ajustar el modelo, tal vez no sea la más adecuada para este caso en particular.

5.2.3. Aplicación del modelo de red neuronal funcional.

La otra metodología no paramétrica propuesta fue el uso de una red neuronal funcional de una sola capa. Al igual que con el caso de la regresión no paramétrica el no suponer linealidad nos da más flexibilidad en el modelo. Sin embargo, los modelos de redes neuronales pueden llegar a tener serios problemas de overfitting. Como se mencionó en el capítulo tres utilizaremos el descenso del gradiente para entrenar el modelo utilizando como elementos de validación el set de datos destinados a test.

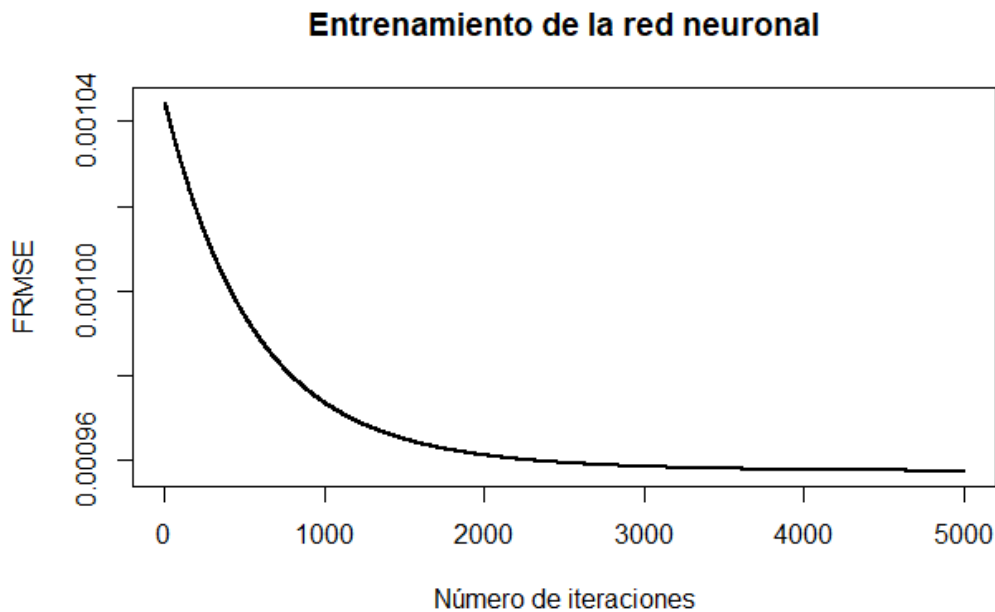


Figura 5.4: Entrenamiento red neuronal.

En la imagen 5.4 podemos ver el comportamiento del entrenamiento de nuestra red. Si aumentamos el número de iteraciones el ajuste del modelo mejorara, pero como podemos ver las mejoras son cada vez más reducidas. En este caso se ha utilizado un learning rate fijo, aunque se podrían utilizar técnicas en las que el parámetro se fuera modificando con el objetivo de obtener una convergencia más rápida.

A diferencia de la regresión no paramétrica este si parece que realizada un buen ajuste en prácticamente toda la serie de tiempo (Imagen 5.5). Parecido a como se comportaba el modelo de regresión paramétrica. Vemos que la traza azul se ajusta muy bien a la curva original, igual que ocurre con la traza roja, lo que nos indica que el modelo no ha tenido ningún problema de overfitting. Una de las cuestiones más particulares que tiene el modelo de redes neuronales el tipo de predicciones que elabora. Debido a que los datos pueden ser bastante abruptos y al tipo de predicción a un paso

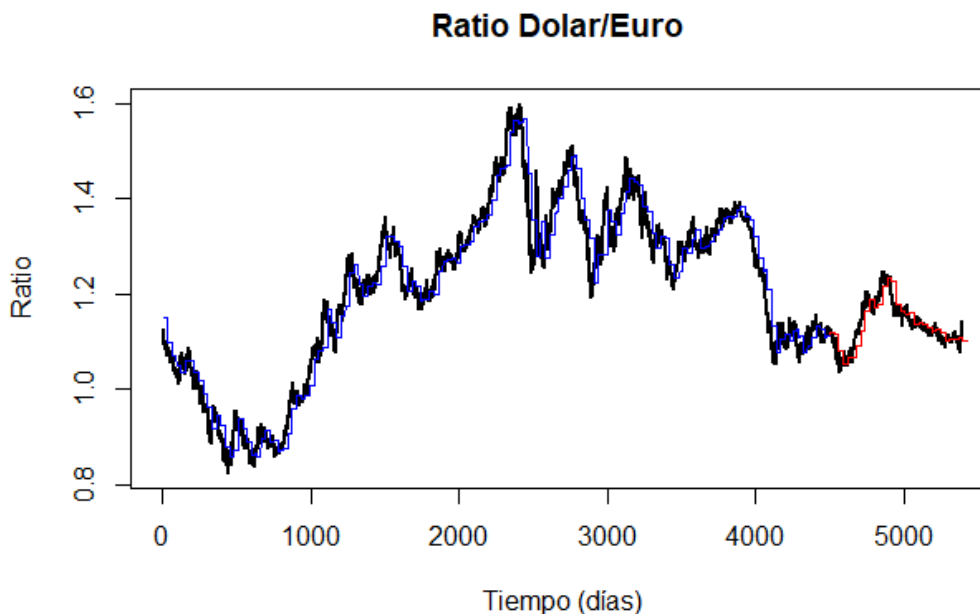


Figura 5.5: Estimación de la red neuronal.

que estamos estudiando, en muchos casos los modelos sufren una separación grande entre el final de una predicción y el inicio de la siguiente. En el caso del modelo de redes neuronales este comportamiento es más claro que en ningún otro modelo.

5.2.4. Aplicación del modelo autorregresivo en espacios de Hilbert.

El último modelo que aplicaremos es el autorregresivo en espacios de Hilbert que fue presentado en el cuarto capítulo. Nos centraremos única y exclusivamente en el modelo autorregresivo de orden uno. Se mencionaron dos tipos distintos de técnicas a la hora de utilizar una base ortonormal para calcular un estimador del operado ρ , la de componentes principales y la base generada por mínimos cuadrados parciales. Para este caso vamos a utilizar componentes principales que es la más común.

Como podemos apreciar en la imagen 5.6 el resultado bastante bueno. Si es cierto que parece que existe un pequeño desfase entre la predicción y el set de datos original, aunque ya es algo que venimos viendo ligeramente en los otros modelos, parece que en este caso es más claro. Cabe destacar que únicamente hemos necesitado una única componente principal.

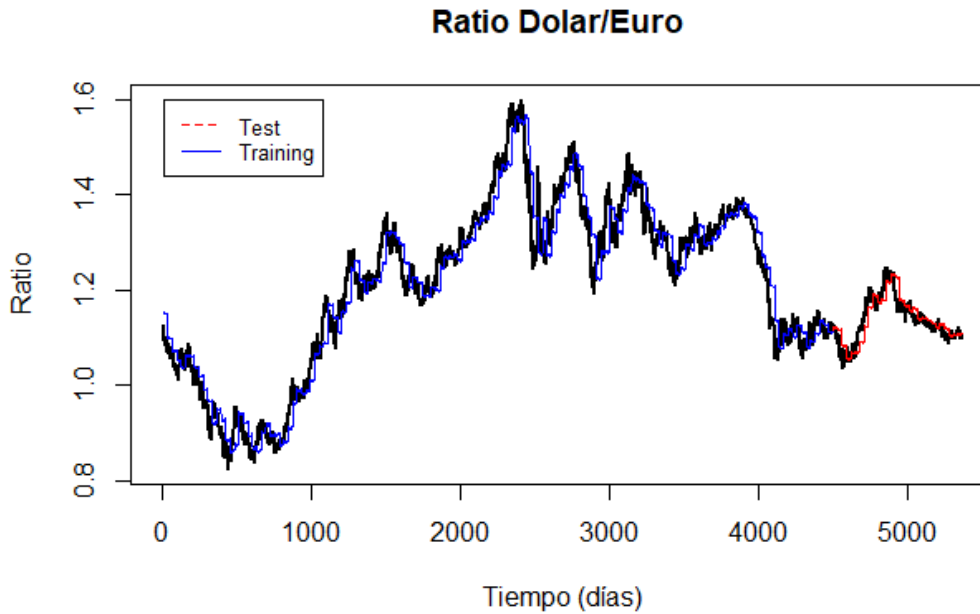


Figura 5.6: Estimación del modelo autorregresivo en espacios de Hilbert.

5.2.5. Comparativa mediante medidas de error.

El determinar cuál de los modelos obtiene un mejor desempeño es algo que difícilmente se puede hacer a simple vista. Es por este motivo que al principio de este tema hemos introducido tres tipos distintos de medidas de error. Nuestro objetivo ahora es utilizar estas medidas con la finalidad de determinar cuál es el modelo más adecuado de todos. En la siguiente tabla se muestran las distintas medidas de error para los distintos modelos.

Modelo\Medida Error	FMAE	FMSE	FRMSE
Paramétrico	0.013	0.0003	0.018
No paramétrico	0.042	0.0023	0.048
Red neuronal	0.016	0.0005	0.022
ARH(1)	0.017	0.0005	0.022

Si bien es cierto que a principios de esta sección se menciona qué es difícil determinar que modelo da mejores resultados a simple vista, el que el modelo paramétrico

de mejores resultados no es algo sorprendente. De la misma forma, que el modelo no paramétrico de peores resultados tampoco es ninguna sorpresa. El debate surge entre los otros dos modelos, el de red neuronal y el ARH(1), y lo cierto es que ambos modelos dan resultados bastante similares, siendo el modelo de redes neuronales ligeramente mejor.

Tenemos por tanto que el mejor modelo para este análisis es el de regresión paramétrico. La principal desventaja es que requerimos de una penalización y el determinarla puede resultar computacionalmente costoso. A de mencionarse que en nuestro análisis hemos determinado el valor de los parámetros de penalización sin ser concienzudos y el resultado aun así ha sido realmente bueno.

En cuanto al modelo de regresión no paramétrica es el que peor resultados obtiene. Esto quizás se deba a que el modelo es considerablemente sensible al uso de una ventana dada, o que tal vez simplemente no sea un modelo adecuado para este caso en particular.

El modelo de redes neuronales si bien es cierto que obtiene los segundos mejores resultados, está muy próximo a el modelo ARH(1). Tiene la ventaja de que con un mayor tiempo de computación se podrían obtener mejores resultados. Sin embargo, su ventaja es a la vez una desventaja pues requiere un tiempo muy grande en comparación con el resto de los modelos para poder obtener un buen resultado.

El último modelo a analizar ha sido el autorregresivo en espacios de Hilbert de orden uno. Con resultados ligeramente peores que los del modelo de redes neuronales, es con diferencia el modelo más rápido de todos. Cabe destacar que únicamente ha necesitado una sola componente principal.

5.3. Aplicación del modelo con respuesta escalar.

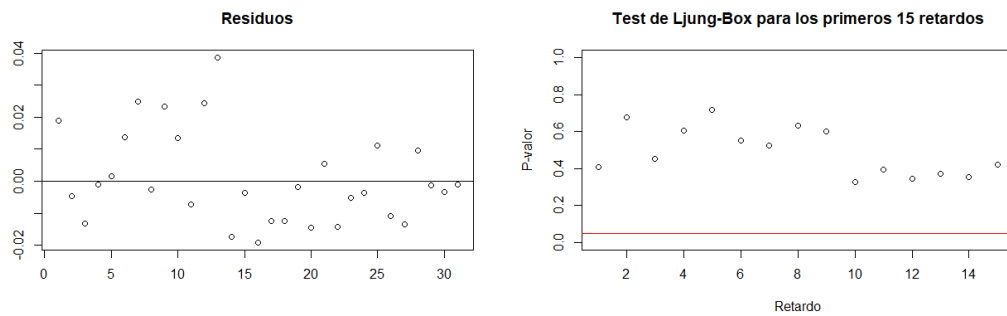
Hasta ahora únicamente nos hemos centrado en analizar el set de datos del ratio entre el precio del dólar y el euro desde un punto vista en el que nuestra respuesta es enteramente funcional. Ahora trataremos de estudiar el mismo problema utilizando el otro enfoque descrito en el trabajo, cuando la respuesta es escalar.

Para enfocar el problema al caso de respuesta escalar hemos de tener propiamente una variable respuesta que sea un escalar. Trabajaremos con una nueva variable que no será más que la media del ratio. El objetivo es por tanto determinar el valor medio del mes siguiente a partir de cómo ha ido evolucionando a lo largo del mes anterior. Puesto que el modelo que dio un mejor resultado en el caso con respuesta funcional fue el paramétrico, solamente nos centraremos en este. Utilizaremos un modelo penalizado idéntico al caso anterior, pero para un modelo con respuesta escalar y otro modelo generado por mínimos cuadrados parciales, ambos descritos en el tema dos. Comenzaremos

estudiando el modelo paramétrico penalizado.

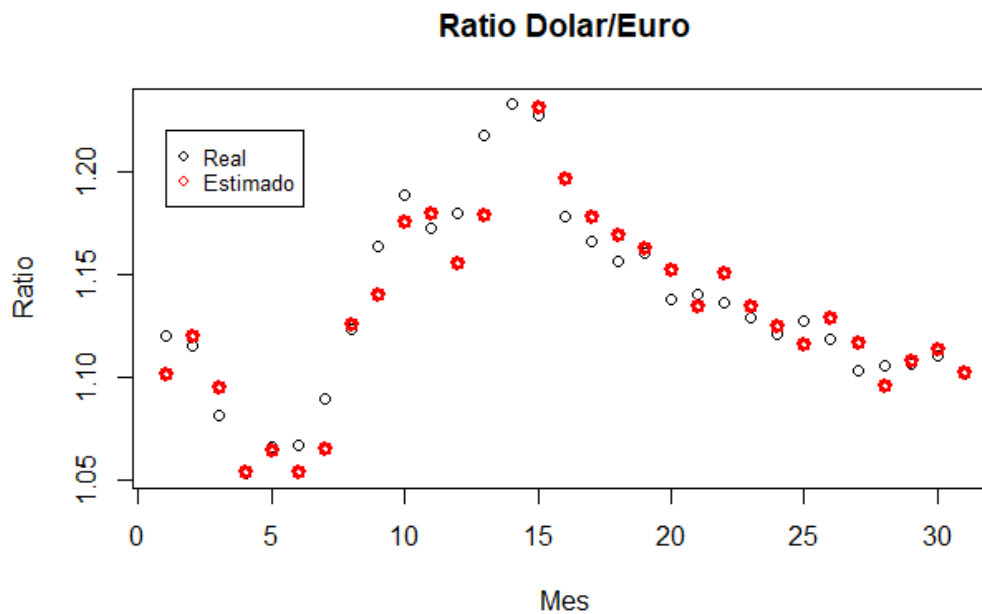
5.3.1. Modelo paramétrico penalizado.

Al igual que hicimos en el caso con respuesta funcional fraccionaremos la serie de tiempo. Lo haremos de una forma idéntica a los apartados anteriores, con la diferencia de que en este caso únicamente nos centraremos en la predicción de los datos no utilizados en el set de entrenamiento.



(a) Residuos

(b) Test Ljung-Box



(c) Estimación modelo paramétrico penalizado

Figura 5.7: Estimación modelo parametrico penalizado

Si nos fijamos en la imagen 5.7(c) tenemos representado en negro los valores reales

mientras que en rojo están nuestras predicciones. Los resultados a primera vista parecen considerablemente buenos. En la imagen 5.7(a) se muestra la gráfica de los residuos del modelo. El comportamiento es cuanto menos un poco sospecho. Para determinar la independencia de los residuos se ha realizado un test de Ljung-Box para los primeros quince retardos. En la imagen 5.7(b) se muestran los p-valores para los distintos test, donde la línea roja representa el $p\text{-valor}=0.05$.

Podríamos realizar un test de normalidad con la esperanza de poder determinar un intervalo de predicción, sin embargo los residuos no son homocedasticos con lo que no podemos utilizar esta vía. Lo que si podemos realizar es wild bootstrap para determinar los intervalos de predicción. Los intervalos de predicción han sido elegidos para alcanzar un 99 % de confianza.

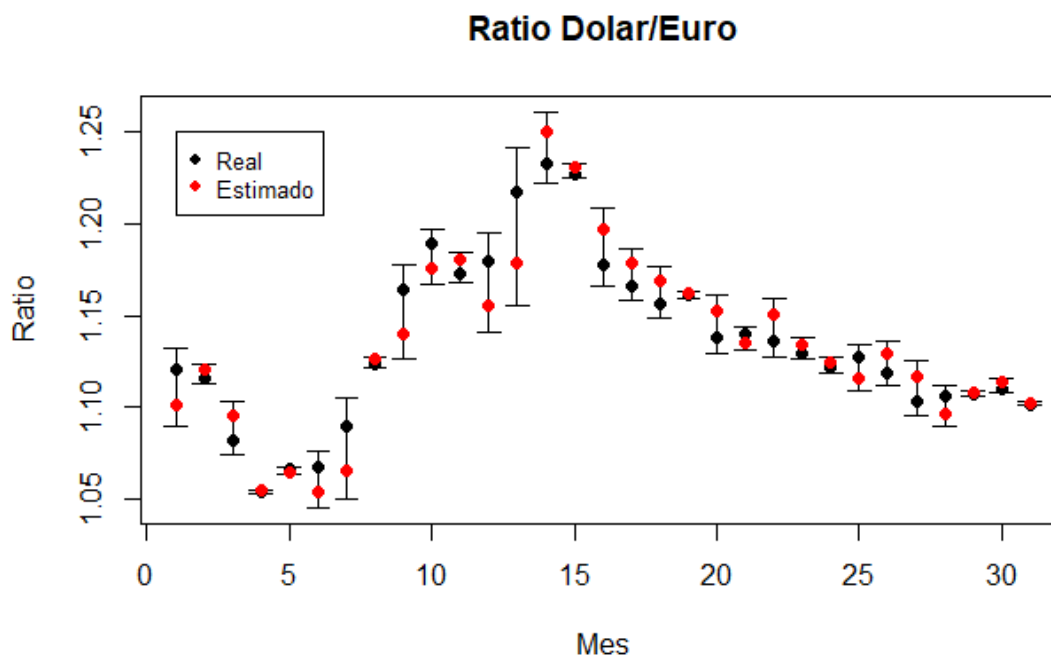
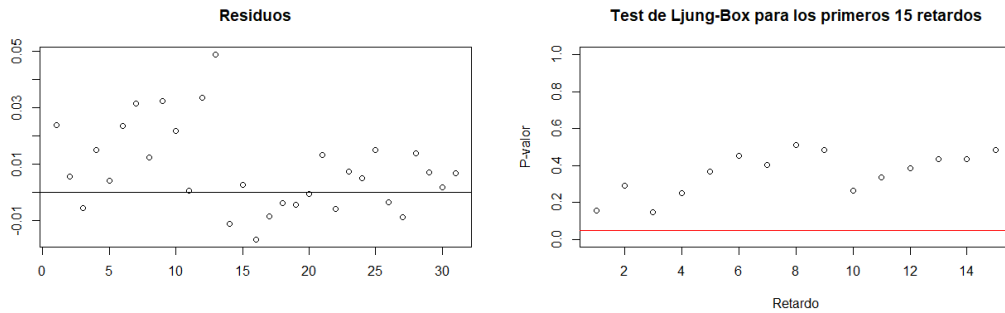


Figura 5.8: Estimación modelo paramétrico penalizado

5.3.2. Modelo de mínimos cuadrados parciales.

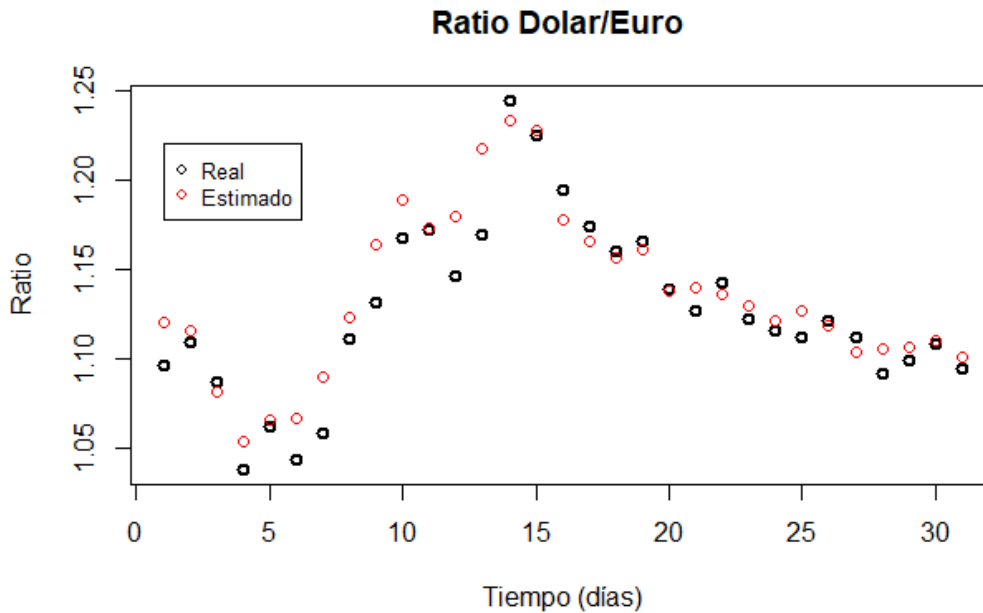
Haremos un análisis idéntico al del modelo paramétrico penalizado con la única diferencia de que en este caso no utilizaremos ninguna penalización. Lo que si tenemos que tener en cuenta de este modelo, es que se centra en determinar una base ortonormal enfocada en conseguir la mejor predicción posible. Una vez que hemos determinado la base, utilizar un mayor número de elementos no implica un mejor resultado. Nos ocurre

por tanto algo similar a los modelos ARH(1) cuando usamos componentes principales. Utilizaremos el error cuadrático medio para determinar el número de elementos de la base más adecuado.



(a) Residuos

(b) Test Ljung-Box



(c) Estimación modelo de mínimos cuadrados parciales

Figura 5.9: Estimación modelo de mínimos cuadrados parciales

Hemos utilizado las tres primeras componentes pls para el modelo. Los resultados son bastante similares a los que obtuvimos en el modelo paramétrico penalizado. Al igual que hicimos en el caso anterior vamos a realizar wild bootstrap sobre los residuos para determinar un intervalo de confianza sobre las predicciones.

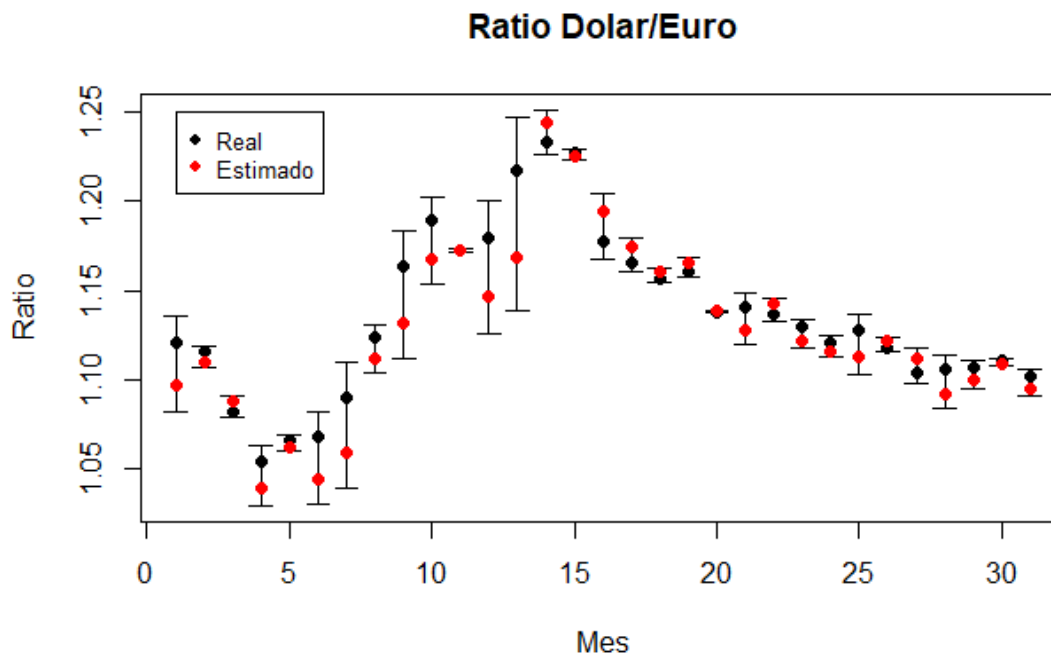


Figura 5.10: Estimación de mínimos cuadrados

5.3.3. Comparativa mediante medidas de error.

Los resultados de ambos modelos son extremadamente similares. Para poder determinar cuál de los dos da un mejor resultado vamos a realizar un análisis similar al hecho en la sección anterior. En este caso trabajaremos con las mismas medidas de error, pero aplicadas al caso no funcional.

Modelo\Medida Error	MAE	MSE	RMSE
Paramétrico	0.011	0.0002	0.014
PLS	0.013	0.0003	0.017

Al igual que paso con el modelo con respuesta funcional el que obtiene mejores resultados es el paramétrico penalizado, aunque los resultados son bastante similares. En ambos casos es necesario calcular el número de componentes a utilizar o penalizaciones, lo que hace que se prácticamente similar el uso de uno u otro, en términos de coste computacional.

Capítulo 6

Conclusiones y futuros trabajos.

Como hemos podido comprobar a lo largo de este trabajo, el análisis de series de tiempo funcionales consigue dar una predicción considerablemente buena. En particular, la aplicación al precio del intercambio del dólar y el euro ha dejado patente el potencial de estas metodologías en un problema con una gran relevancia para el sector financiero.

De las dos visiones tratadas en el trabajo, la completamente funcional es posiblemente la menos común en la literatura (aunque cada vez más frecuente). Si bien es cierto que el modelo con mejores resultados ha sido el paramétrico penalizado, cabe destacar el modelo funcional de red neuronal autorregresiva, pues es una visión completamente nueva y tal vez en otro tipo de análisis pueda dar un mejor resultado.

En cuanto al estudio con respuesta escalar el modelo paramétrico penalizado vuelve a dar mejores resultados, lo cual es algo curioso pues el enfoque de los mínimos cuadrados parciales es encontrar aquellas componentes que minimicen el error de predicción. De todas formas, en ambos casos los resultados han sido muy buenos y bastante próximos entre sí.

Si nos fijamos en el modelo paramétrico ordinario con respuesta funcional presentado en el capítulo dos y el autorregresivo en espacios de Hilbert de orden uno, se aprecia claramente que son distintas estimaciones de un mismo modelo. Es algo que ya mencionamos en el capítulo dos y que ahora cabe resaltar, pues en nuestro análisis final solo hay una diferencia entre ellos, las penalizaciones. Como pudimos comprobar el resultado mejora con estas penalizaciones. Lo cierto es que el hecho de que en ambos casos el modelo penalizado de mejores resultados, nos deja claro que el uso de penalizaciones ha de ser una técnica muy tenida en cuenta.

El único modelo que ha dado unos resultados ligeramente peores ha sido el no paramétrico. Aunque como ya mencionamos en el capítulo tres, los modelos de regresión no paramétrica funcional pueden ser bastantes sensibles a la ventana elegida. El realizar

un trabajo más exhaustivo de este enfoque implicaría estudiar distintos núcleos, técnicas de determinación de la ventana o incluso trabajar con distintas métricas. Todo esto requeriría hacer un trabajo en si solo dedicado al modelo no paramétrico. Por tanto en este caso no debemos determinar que el modelo no es adecuado, sino que requiere un estudio más profundo.

Otra de las cuestiones en las que no hemos entrado en mucho detalle en este trabajo, es en el uso del wild bootstrap para la regresión funcional con respuesta escalar. Un poco parecido a lo que mencionamos con el modelo de regresión no paramétrica, el realizar un análisis sobre las técnicas bootstrap a la regresión funcional podría ser un tanto excesivo, teniendo en cuenta todo lo ya realizado. Mencionar que en este trabajo únicamente se han utilizado las ideas propuestas por Ferraty et al.(2010)[11].

Tras este análisis existen distintas vías más avanzadas que aquí no se han tratado y que pueden ser bastante interesantes. Entre ellas destacaremos:

- El uso de componentes PLS como base ortonormal en los modelos autorregresivos en espacios de Hilbert.
- El uso de penalizaciones en los mínimos cuadrados parciales.
- Estudiar la regresión con respuesta funcional utilizando los mínimos cuadrados parciales.
- Un estudio más profundo de los modelos no parametricos.
- Los modelos autorregresivos en espacios de Hilbert de orden p .
- Los modelos autorregresivos en espacios de Hilbert con variables exógenas.
- Los modelos autorregresivos de media móvil en espacios de Hilbert.

Apéndice A

Demostraciones importantes.

A.1. Demostración Lema 4.1.

En este primer apéndice daremos una breve demostración de uno de los lemas más importantes en el estudio de la existencia de los modelos autorregresivos en espacios de Hilbert de orden 1. Comenzaremos reescribiendo de nuevo el lema.

LEMA 4.1. Para todo $\rho \in \mathcal{L}$ las siguientes dos condiciones son equivalentes.

- c1) Existe un entero $j_0 \geq 1$ tal que $\|\rho^{j_0}\|_{\mathcal{L}} < 1$
- c2) Existe un $a > 0$ y $0 < b < 1$ tal que $\|\rho^j\|_{\mathcal{L}} < ab^j, j \geq 0$

Demostración. La condición c1 implica c0 es simple de demostrar, únicamente necesitamos imponer la condición de que $ab^j = 1$ para todo $j > 0$. Que c0 implica c1 no es una cuestión que se pueda ver a simple vista. Comenzaremos escribiendo j como $j = j_0q + r$ donde $q \geq 0$ y $0 \leq r < j_0$.

$$\|\rho^j\| = \|\rho^{j_0q} \rho^r\| \leq \|\rho^{j_0}\|^q \|\rho^r\|$$

El caso de que $\|\rho^{j_0}\| = 0$ implica que $\|\rho^j\| = 0$, con lo que se cumpliría la condición c1 para todo $0 < a$ y $0 < b < 1$. Pasemos ahora al caso en el que $\|\rho^{j_0}\| > 0$. Puesto que $q > j/j_0 - 1$ y $\|\rho^{j_0}\| < 1$, obtenemos

$$\|\rho^j\| \leq \|\rho^{j_0}\|^{j/j_0-1} \|\rho^r\| \leq (\|\rho^{j_0}\|^{1/j_0})^j \|\rho^{j_0}\|^{-1} \max_{0 \leq r < j_0} \|\rho^r\|$$

Si definimos ahora $a' = \|\rho^{j_0}\|^{-1} \max_{0 \leq r < j_0} \|\rho^r\|$ y $b' = \|\rho^{j_0}\|^{-1}/j_0$ tenemos

$$\|\rho^j\|_{\mathcal{L}} < a'b'^j$$

Puesto que $\|\rho^{j_0}\| < 1$ implica que $0 < b' < 1$ y puesto que $a' = \|\rho^{j_0}\|^{-1}k$ con $k > 0$ implica que $a' > 0$, por lo que cumplen las condiciones impuestas para c1.

A.2. Demostración teorema 4.1.

En este apéndice se dará una demostración de el teorema pilar de este trabajo, que nos determina la existencia del modelo autorregresivo en espacio de Hilbert de orden 1. El teorema dice lo siguiente.

TEOREMA 4.1. Si la condición c1) del lema es cierta, entonces la ecuación 4.1 tiene una única solución estacionaria que viene dada por

$$X_n = \sum_{i=0}^{\infty} \rho^i(\epsilon_{n-i}), \quad n \in \mathbb{Z}$$

La serie converge de forma casi segura en L^2 , es decir

$$E\|X_n - \sum_{i=0}^l \rho^i(\epsilon_{n-i})\| \rightarrow 0, \quad l \rightarrow \infty$$

La demostración ahora propuesta se basa en las ideas de Kokoszka et al.(2012)[14].

Demostración.

A lo largo del trabajo nos hemos centrado especialmente el espacio de Hilbert L^2 , para esta demostración en concreto trabajaremos con el espacio $L^2 = L^2[0, 1]$. Si tenemos funciones aleatorias definidas sobre un espacio de probabilidad Ω , entonces nuestro espacio de Hilbert de trabajo sera $L^2(\Omega, L^2[0, 1])$ que tiene como producto interno $E\langle X, Y \rangle$, $X, Y \in L^2[0, 1]$. Para determinar que la sucesión $X_n^m = \sum_{i=0}^m \rho^i(\epsilon_{n-i})$ tiene límite en $L^2(\Omega, L^2[0, 1])$, es suficiente con demostrar que es una sucesión de Cauchy para cada n fijo.

$$E\left\|\sum_{i=m}^{m'} \rho^i(\epsilon_{n-i})\right\|^2 = \sum_{i=m}^{m'} \sum_{k=m}^{m'} E\langle \rho^i(\epsilon_{n-i}), \rho^k(\epsilon_{n-k}) \rangle = E\sum_{i=m}^{m'} \|\rho^i(\epsilon_{n-i})\|^2$$

Como se mencionó en el capítulo sobre conceptos de datos funcionales la esperanza conmuta con los operadores $\rho \in \mathcal{L}$, por lo tanto $E\rho^j(\epsilon_{n-i}) = 0$. Por el lema 4.1

$$E\left\|\sum_{i=m}^{m'} \rho^i(\epsilon_{n-i})\right\|^2 \leq \left(\sum_{i=m}^{m'} \|\rho^i\|^2\right) E\|\epsilon_0\|^2 \leq E\|\epsilon_0\|^2 a^2 \sum_{j=m}^{m'} b^{2j}$$

con lo que $X_n^{(m)}$ converge en $L^2(\Omega, L^2[0, 1])$.

Una vez demostrado que la sucesión tiene límite en $L^2(\Omega, L^2[0, 1])$ vamos a estudiar su convergencia. Para determinar la convergencia primero verificaremos que

$$\sum_{i=m}^{m'} \|\rho^i(\epsilon_{n-i})\| < \infty \quad c.s.$$

Esto se cumple por la condición c1 del lema 4.1

$$E\left(\sum_{i=0}^{\infty} \|\rho^i\| \|\epsilon_{n-i}\|\right)^2 \leq \sum_{i,k=0}^{\infty} \|\rho^i\| \|\rho^k\| E\|\epsilon_0\|^2 \leq E\|\epsilon_0\|^2 \left(\sum_{i=0}^{\infty} ab^i\right)^2 < \infty$$

Supongamos que tenemos $\{X'_n\}$ una sucesión que satisface la ecuación 4.1. Podemos escribir la sucesión de la siguiente forma

$$X'_n = \sum_{i=1}^m \rho^i(\epsilon_{n-i}) + \rho^{m+1}(X'_{n-m+1})$$

donde $m \geq 1$. Por tanto,

$$E\|X'_n - X_n^{(m)}\| \leq \|\rho^{m+1}\| E\|X'_{n-m+1}\| \leq E\|X_0\| ab^{m+1}$$

Así X'_n es igual de forma casi segura al límite de $X_n^{(m)}$ es decir a X_n .

Apéndice B

Estudio detallado de los datos.

El apartado final de este trabajo se centra en el análisis de las distintas metodologías explicadas a lo largo del tema. Tanto si se trata de un enfoque con respuesta funcional o escalar se ha hecho el análisis entorno al set de datos del precio del intercambio entre el dólar y el euro. El set de datos utilizado ha sido recogido de la base de datos del banco central europeo donde se recogen todos los datos desde la formación de la moneda en enero de 1999.

Cuando trabajamos con datos funcionales es común que nuestros datos tengan una estructura bien definida. Sin embargo, cuando tratamos de estudiar las series de tiempo desde un enfoque funcional uno de los principales problemas que presentan es que no tienen esta estructura bien definida, forman parte de un continuo. Por este motivo el punto de partida es dividir nuestra serie de tiempo en intervalos de igual tamaño, donde cada una de estas secciones formará un dato funcional. Dado que nuestra serie de tiempo recoge un único valor diario, nuestra unidad mínima de medida será por tanto un día. Hemos de determinar ahora cuantos días hemos de utilizar para formar un dato funcional. Quizás de forma intuitiva uno podría suponer que trabajar con datos recogidos semana a semana es una buena idea. El principal problema que tiene este enfoque es solo se recogen cinco datos a la semana, lo cual supone un número muy pequeño de puntos a la hora de formar nuestro dato funcional. Por este motivo nos centraremos en tomar medidas de tiempo que podría considerarse como "mensuales". Si bien es cierto que un mes son entorno a veintitrés días laborables, nosotros por simplicidad trabajaremos con datos formados por treinta días.

Una vez que hemos determinado como particionar nuestra serie de tiempo, el siguiente punto es elegir una base en la que representaremos nuestro set de datos funcionales. En este trabajo se han explicado tres de los métodos más comunes a la hora de representar los datos funcionales, aunque en este caso únicamente se utilizara la base de Bsplines. Para determinar el número de elementos más adecuado para la representación se ha utilizado la función `optim.basis` del paquete `fda.usc` [8], que calcula cual es el número de elementos más adecuados por validación cruzada o por validación cruzada

generalizada. En toda esta explicación se está hablando de la representación de los datos funcionales, esto no quiere decir que no hayamos usado las componentes principales o los mínimos cuadrados parciales para calcular un modelo en particular.

Apéndice C

Código.

C.1. Código modelo paramétrico.

```
#Libreria principal del analisis de datos funcionales.
library(fda.usc)

#Base optima
opt.bsY<-optim.basis(fdata(t(y)),numbasis = seq(4,30,2),type.basis = "bspline")
opt.bsX<-optim.basis(fdata(t(x)),numbasis = seq(4,30,2),type.basis = "bspline")

#Número de elementos de la base óptima
nbx<-opt.bsX$numbasis.opt
nby<-opt.bsY$numbasis.opt

#Creamos las bases y formamos nuestro set de datos funcionales.
basisx <- create.bspline.basis(c(0, 1),nbasis = nbx,norder = 5 )
basisy<- create.bspline.basis(c(0,1),nbasis = nby,norder = 5 )
datosf<-Data2fd((x),seq(0,1,length=30),basisobj = basisx)
datosfN<-Data2fd((y),seq(0,1,length=30),basisobj = basisy)

#Inicializacion de los parametros de penalización
lambdas<-1
lambdat<-1

#Estimación rápida del orden de la penalización
FMSE<-0
\textcolor{0orange}{for} (i in 1:100) {
  lambdas<-lambdas/10
  lambdat<-lambdat/10
  res.fr = fregre.basis.fr(datosf[1:150],datosfN[1:150],basisx,basisy,
```

```

lambda.s = lambdas,lambda.t = lambdat,Lfdobj.s = 1,Lfdobj.t = 1)
prf<-predict(res.fr,datosf[151:179],basisobj = basisx)
pr<-eval.fd(prf,seq(0,1,length=ttiempo))
FMSE[i]<-mean(colMeans((y[,151:179]-pr)^2))
}

#FMSE mínimo
which.min(FMSE)

#Modelo final

lambdas<-1/10^which.min(FMSE)
lambdat<-1/10^which.min(FMSE)
res.fr = fregre.basis.fr(datosf[1:150],datosfN[1:150],basisx,basisy,
lambda.s = lambdas,lambda.t = lambdat,Lfdobj.s = 1,Lfdobj.t = 1)
prf<-predict(res.fr,datosf,basisobj = basisx)
pr<-eval.fd(prf,seq(0,1,length=ttiempo))

#Representación Grafica
plot.ts(as.vector(datos[1:(length(datos)-60)]),lwd=2,lty=1,
main="Ratio Dolar/Euro ",xlab="Tiempo (días)", ylab="Ratio")
lines(as.vector(pr),col="red")
lines(as.vector(pr[,1:150]),col="blue")
legend(1,1.6 , legend=c("Test", "Training"),
      col=c("red", "blue"), lty=2:1, cex=0.8)

```

C.2. código modelo no paramétrico.

```

#Librerias
library(fda.usc)

#Base optima
opt.bsY<-optim.basis(fdata(t(y)),numbasis = seq(4,30,2),type.basis = "bspline")
opt.bsX<-optim.basis(fdata(t(x)),numbasis = seq(4,30,2),type.basis = "bspline")

#Número de elementos de la base óptima
nbx<-opt.bsX$numbasis.opt
nby<-opt.bsY$numbasis.opt

#Creamos las bases y formamos nuestro set de datos funcionales.
basisx <- create.bspline.basis(c(0, 1),nbasis = nbx,norder = 5 )

```



```

basisy<- create.bspline.basis(c(0,1),nbasis = nby,norder = 5 )
X<-Data2fd((x),seq(0,1,length=30),basisobj = basisx)
Y<-Data2fd((y),seq(0,1,length=30),basisobj = basisy)

```

```

#Nucleo Gaussiano

```

```

W <- function(d, h) {
  k <- (1 / (2 * pi) ^ (1 / 2)) * exp(-((d) / h) ^ 2 / 2)
  return(k)
}

```

```

#Regresion noparametrica funcional

```

```

NPregreSo <- function(X,Y1,h) {
  d<-0
  K<-0
  n<-ncol(X$coefs)
  for (i in 1:n) {
    d[i]<-mean((Y1$coefs-X[i]$coefs)^2)^1/2
  }
  for (i in 1:n) {
    K[i]<-W(d[i],h)
  }

  w<-K/sum(K)

  X1<-matrix(0,ncol = 1,nrow = nrow(X$coefs))
  for (i in 1:(n-1)) {
    X1<-X1+w[i]*X[i+1]$coefs
  }
  return(X1)
}

```

```

#Determinación de la ventana por CV

```

```

h<-0.01
CVf<-0
for (j in 1:10) {
  CV<-0
  h<-h/1.5
  print(h)
  for (i in 1:179) {
    npr<-NPregreSo(X[-i],X[i],h)
    CV[i]<-(colMeans((Y[i]$coefs-npr)^2))
  }
}

```

```

    }
    CVf[j]<-mean(CV)
    print(j)
  }

#Modelo final no parametrico
h<-0.01/(1.5^which.min(CVf))
nprf<-matrix(0,nrow = 30,ncol = 179)
for (i in 1:179) {
  npr<-NPregreSo(X[1:50],X[i],h)
  nprf[,i]<-eval.fd(Data2fd(npr),seq(0,1,length=ttiempo))
}

#Representación Grafica
plot.ts(as.vector(datos[1:(length(datos)-60)]),lwd=2,lty=1,
main="Ratio Dolar/Euro ",xlab="Tiempo (días)", ylab="Ratio")
lines(as.vector(nprf),col="red")
lines(as.vector(nprf[,1:150]),col="blue")
legend(1,1.6 , legend=c("Test", "Training"),
      col=c("red", "blue"), lty=2:1, cex=0.8)

```

C.3. Código del modelo red neuronal.

```

#Librerias
library(fda.usc)
library(matrixcalc)

# Función de activación Sigmoide
sigmoid <- function(x) return(1/(1+exp(-x)))

#Derivada de la función sigmoide
d.sigmoid <- function(x) return(exp(-x)/(1+exp(-x))^2)

#Semilla
set.seed(12345)

#Inicialización de los pesos
hl<-5
W1 <- matrix(runif(4*hl[1], 0, 0.1), nrow = 4)
W2<-W1
#Discretización de la componente temporal

```

```

s<-seq(0,1,length.out = 30)
t<-seq(0,1,length.out = 30)

u<-0

#Ratio de aprendizaje
rate<-1E-2
X<-(serie.r)

#Red neuronal
for (l in 1:5000) {

#Matriz de pesos
P<-matrix(0,nrow = 30,ncol = 30)
Pf<-matrix(0,nrow = 30,ncol = 30)
for (k in 1:h1) {
  for (i in 1:30) {
    for (j in 1:30) {
      P[i,j]<-W1[4,k]*sigmoid(W1[1,k]+W1[2,k]*s[i]+W1[3,k]*t[j])
    }
  }
  Pf<-Pf+P
}

#Estimacion
yest<-t(Pf)%*%X[,1:150]/(30)

#Calculo FMRSE
er<-0
for (i in 1:150) {
  er[i]<-t(X[,i+1]-yest[,i])%*(X[,i+1]-yest[,i])/30
}
u[l]<-mean(er)
if (u[l]==u[which.min(u)]) {
  W3<-W1
}

#Dencenso del gradiente
Pd<-matrix(0,nrow = 30,ncol = 30)
Pdf<-list(Pd)
for (k in 1:h1) {
  for (i in 1:30) {
    for (j in 1:30) {

```

```

        Pd[i,j]<-W1[4,k]*d.sigmoid(W1[1,k]+W1[2,k]*s[i]+W1[3,k]*t[j])
    }
}
Pdf[[k]]<-Pd
}
STB<-matrix(1,nrow = 30,ncol = 4)
STB[,2]<-s
STB[,3]<-t
grad<-matrix(ncol = h1,nrow = 4)
L<-0
for (k in 1:h1) {
  for (j in 1:4) {
    for (i in 1:150) {
      if (j==1) {
        L[i]<-t((X[,i+1]-yest[,i]))**Pdf[[k]]**
          hadamard.prod(STB[,j],X[,i+1])/150^2
      }
      if (j==2) {
        L[i]<-t((X[,i+1]-yest[,i]))**Pdf[[k]]**
          hadamard.prod(STB[,j],X[,i+1])/150^2
      }
      if (j==3) {
        L[i]<-t(hadamard.prod((X[,i+1]-yest[,i]),STB[,j]))**
          Pdf[[k]]**X[,i+1]/150^2
      }
      if (j==4) {
        L[i]<-t((X[,i+1]-yest[,i]))**Pf**
          hadamard.prod(STB[,j],X[,i+1])/150^2}
      if (j==5) {
        L[i]<-mean(X[,i+1]-yest[,i])*mean(X[,i+1])}
    }
    grad[j,k]<-sum(L)
  }
}
W1<-W1+rate*grad
if (1 %% 100==0) {
  plot(u)
}

#Plot training
plot.ts(as.vector(u[1:5000]),lwd=2,lty=1,
main="Entrenamiento de la red neuronal",xlab="Número de iteraciones", ylab="FRMSE")

#Representación Grafica

```

```

yest<-t(Pf)*%X/(30)
plot.ts(as.vector(datos[1:(length(datos)-60)]),
lwd=2,lty=1,main="Ratio Dolar/Euro ",xlab="Tiempo (días)", ylab="Ratio")
lines(as.vector(yest),col="red")
lines(as.vector(yest[,1:150]),col="blue")
legend(1,1.6 , legend=c("Test", "Training"),
      col=c("red", "blue"), lty=2:1, cex=0.8)

```

C.4. Código modelo ARH(1).

C.4.1. Función ARH(1).

```

#Modelo autorregresivo en espacios de Hilbert de orden 1
#serie.f= set de datos en forma de vector
#t= número de elementos del dato funcional
#pc= número de componentes principales a calcular
#ttraining= número de componentes del set de datos

ARH1N<-function(serie.f,t,pc,ttraining){
  m<-t-1

  #Reescribir los datos en forma de matriz
  serie<-t(matrix(serie.f,nrow=t))

  #Eliminamos los elementos repetidos de la matriz
  #que se generan cuando t no es multiplo del número de datos
  N1<-2
  if (length(serie.f)%t==0) {
    N1<-1
  }

  #Generamos nuestros datos funcionales
  fdx<-Data2fd(t(serie[1:(nrow(serie)-N1+1),]),argvals = 0:m/m)

  #Determinamos las componentes principales
  fdXpca=pca.fd(fdx,pc)
  eigenvalues=fdXpca$values; scoresX=fdXpca$scores
  harmonicsX=fdXpca$harmonics
  N<-nrow(serie)-N1
  media<-mean.fd(fdx[1:N])

```

```

#Calculamos el número de componentes principales que
#explican el 0.99% de la varianza
l<-1
for (i in 1:pc) {
  l<-1-fdXpca$varprop[i]
  if (l<0.01) {
    nbasis<-i
    break()
  }
}
print(nbasis)

#Función nucleo
rho1<-Nucleophi(fdx[1:ttraing],eigenvalues,scoresX,harmonicsX,m,nbasis,N)

#Guardamos nuestros datos funcionales en forma de matriz
xn<-matrix(0,nrow = t,ncol = N)
for (l in 2:N) {
  XN<-Data2fd(rho1%*%eval.fd(evalarg = 0:m/m,fdx[l]-media)/t,argvals = 0:m/m)+media
  xn[,l]<-eval.fd(XN,seq(0,1,length=t))
}
return(xn)
}

#Representación Grafica
plot.ts(as.vector(datos[30:(length(datos)-60)]),lwd=2,lty=1,
main="Ratio Dolar/Euro ",xlab="Tiempo (días)", ylab="Ratio")
lines(as.vector(xn[,2:180]),col="red")
lines(as.vector(xn[,2:150]),col="blue")
legend(1,1.6 , legend=c("Test", "Training"),
      col=c("red", "blue"), lty=2:1, cex=0.8)

```

C.4.2. Función Nucleophi.

```

Nucleophi<-function(fdx,eigenvalues,scoresX,harmonicsX,m,base,N){
  rho<-0
  rho1<-0
  for (l in 1:base) {
    for (i in 1:(N-1)) {
      for (j in 1:base) {

```

```
        rho<-as.numeric((1/eigenvalues[j])*scoresX[i,j]*
        scoresX[i+1,1])*eval.fd(evalarg=0:m/m, harmonicsX[j])%*%
        t(eval.fd(evalarg=0:m/m, harmonicsX[1]))
        rho1<-rho+rho1
    }
}
print(1)
}
rho1<-rho1/(N-1)
return(rho1)
}
```


Bibliografía

- [1] Aneiros-Pérez G, Cao R, Vilar-Fernández JM (2011) Functional methods for time series prediction: A nonparametric approach. *Journal of Forecasting* 30:377-392.
- [2] Aneiros-Pérez G, Cao R, Vilar-Fernández JM (2013) Functional Prediction for the Residual Demand in Electricity Spot Markets . *IEEE Transactions on Power Systems*, volumen 28.
- [3] Beyaztas U, Shang HL (2020) On function-on-function regression: partial least squares approach. *Environ Ecol Stat* 27:95-114.
- [4] Boor C (1978) *A Practical Guide to Splines*. Springer.
- [5] Bosq D (2000) *Linear Processes in Function Spaces*. Springer.
- [6] Damon J, Guillas S (2005) Estimation and Simulation of Autoregressive Hilbertian Processes with Exogenous Variables. *Statistical Inference for Stochastic Processes* 8: 185?204.
- [7] Damon J, Guillas S (2002) The inclusion of exogenous variables in functional autoregressive ozone forecasting. *Environmetrics* 8: 13: 759?774.
- [8] Febrero-Bande M, Oviedo de la Fuente M (2012) Statistical Computing in Functional Data Analysis: The R Package *fda.usc*. *J Stat Softw*.
- [9] Febrero-Bande M, Galeano P, Gozalez-Manteiga W (2017) Functional principal component regression and functional partial least-squares regression: An overview and a comparative study. *International Statistical Review* 85:61-83.
- [10] Ferraty F, Mas A, Vieu P (2016) Nonparametric regression on functional data: inference and practical aspects. *Aust. N.Z.J. Stat.* 49:267-286.
- [11] Ferraty F, Keilegom I, Vieu P (2010) On the Validity of the Bootstrap in Non-Parametric Functional Regression. *Scandinavian Journal of Statistics.* 37: 286-306.
- [12] Ferraty F, Vieu P (2006) *Nonparametric Functional Data Analysis*. Springer.
- [13] Hyndman RJ, Athanasopoulos G (2018) *Forecasting: Principles and Practice*. OTexts, pp 443-450.

- [14] Kokoszka P, Horváth L (2012) Inference for Functional Data with Applications. Springer.
- [15] Peña D (2010) Análisis de series de temporales. Alianza Editorial
- [16] Portela J (2017) Functional time series forecasting in electricity markets: a novel parametric approach. Tesis, Universidad Pontificia Comillas.
- [17] Preda C, Saporta G (2005) PLS regression on a stochastic process. Computational Statistics & Data Analysis 48:149-158.
- [18] Ramsay JO, Silverman BW (2005) Functional Data Analysis. Springer.