

Università degli Studi di Padova
Dipartimento di Scienze Statistiche
Corso di Laurea Magistrale in
Scienze Statistiche



**MODELLI DI REGRESSIONE FUNZIONALE CON STRUTTURA DI
SPARSITÀ CONDIVISA**

Relatore Prof. Antonio Canale
Dipartimento di Scienze Statistiche

Laureanda: Marta Gaio
Matricola N 2023837

Anno Accademico 2021/2022

Indice

Introduzione	1
1 I dati funzionali	3
1.1 Da dati grezzi a funzioni lisce	3
1.2 Tecniche di lisciamiento	4
1.3 Sistemi di funzioni di base	6
1.3.1 <i>B-splines</i>	6
1.3.2 Basi di Fourier	7
2 Modello di regressione lineare funzionale	9
2.1 Modello di regressione funzionale classico	10
2.1.1 Interpretazione del coefficiente funzionale	11
2.2 Modello di regressione lineare funzionale localmente sparso	12
2.2.1 Rappresentazione in basi del coefficiente funzionale	13
2.2.2 Penalizzazione che induce sparsità	14
2.2.3 Funzione obiettivo	16
3 Modello di regressione lineare funzionale con struttura di sparsità condivisa	19
3.1 Proprietà desiderate	19
3.2 Specificazione del modello	21
3.2.1 Penalizzazione che induce sparsità con <i>pattern</i> condiviso	22
3.3 Controparte empirica delle quantità descritte	24
3.4 Trasformazione delle covariate funzionali	26

4	Studio di simulazione	29
4.1	Parametri dello studio	29
4.2	Modelli alternativi	32
4.3	Risultati	34
5	Analisi di alcuni dati meteorologici	39
5.1	Descrizione del problema e dei dati	39
5.2	Analisi esplorativa e preparazione dei dati	40
5.3	Stima del modello di regressione funzionale con sparsità condivisa	43
5.4	Risultati	45
	Conclusioni	49
A	Strumenti di algebra e analisi matematica	51
A.1	Operatore differenziale	51
A.2	Formula di Cox-De Boor	52
A.3	Serie di Fourier	53
A.4	Prodotto di Hadamard	53
B	Risultati aggiuntivi dello studio di simulazione	55
C	Codice R utilizzato	59
	Bibliografia	65

Introduzione

Nell'analisi statistica supervisionata esistono fondamentalmente due scopi principali: il raggiungimento di una buona accuratezza previsiva e la scoperta di variabili predittive importanti (Zou, 2006). Solo i modelli che tengono conto congiuntamente di accuratezza previsiva e interpretabilità permettono infatti di comprendere il comportamento complessivo del fenomeno in esame o individuare delle particolarità nell'andamento dei dati. Nei modelli di regressione per l'analisi dei dati euclidei *standard* per favorire l'interpretabilità del modello la strada che ha avuto più successo e che è stata maggiormente studiata nella letteratura è quella della selezione delle variabili, che permette di individuare le uniche covariate che sono veramente in relazione con la variabile risposta.

Il concetto di selezione delle variabili può essere in qualche modo trasferito anche nel contesto dell'analisi dei dati funzionali, che è quella branca della statistica che studia la modellazione di dati che giacciono in spazi infinito-dimensionali. Nei modelli di regressione per dati funzionali, per ciascuna variabile esplicativa funzionale viene stimato un coefficiente di regressione che è a sua volta una funzione, ma per gran parte dei fenomeni che si possono analizzare è ragionevole assumere che le covariate abbiano veramente un'influenza sulla variabile risposta solo in certe regioni del loro dominio. Per favorire l'interpretabilità del modello, è stato dunque necessario introdurre degli strumenti modellistici in grado di selezionare tali regioni, incoraggiando una stima verso lo zero dei coefficienti funzionali nelle zone in cui non esiste una relazione tra le covariate e la risposta. Inoltre, qualora siano disponibili delle informazioni aggiuntive riguardo la struttura di sparsità dei coefficienti di regressione funzionali, è utile servirsi di strumenti adeguati in grado

di codificarle. In questo modo l'interpretabilità del modello verrebbe ulteriormente facilitata, potenzialmente riducendo al contempo l'errore di previsione.

In questo elaborato viene proposto un metodo per la regressione funzionale utile in un contesto come quello appena descritto. Nello specifico, viene preso in esame il caso in cui vi siano due covariate funzionali che influenzano la variabile risposta scalare potenzialmente nelle stesse regioni del loro dominio comune. Il metodo proposto prevede l'ottimizzazione di una funzione obiettivo in cui è presente una particolare penalizzazione, che oltre a permette che le stime dei coefficienti funzionali siano sparse, consente che esse abbiano la medesima struttura di sparsità.

Il presente lavoro verrà strutturato come segue. Nel Capitolo 1 verranno delineate le caratteristiche peculiari dei dati funzionali e verrà descritto l'approccio tipicamente usato per la loro rappresentazione, che prevede l'utilizzo di strumenti matematici appropriati, in conseguenza al fatto che i dati analizzati giacciono in uno spazio funzionale e non euclideo. Nel Capitolo 2 sarà riportata una trattazione riguardante uno dei più importanti strumenti dell'analisi dei dati funzionali, ossia il modello di regressione lineare funzionale. In seguito, verrà descritto anche il modello di regressione localmente sparso recentemente proposto da Bernardi et al. (2022), che permette di ottenere una particolare struttura di sparsità del coefficiente funzionale stimato. Nel Capitolo 3, invece, verrà proposta una soluzione al problema oggetto di studio in questo lavoro, introducendo il modello di regressione lineare funzionale con struttura di sparsità condivisa. Per valutare le prestazioni del modello proposto, confrontandole anche con quelle di alcuni modelli concorrenti, nel Capitolo 4 viene poi condotto uno studio di simulazione. Infine, nell'ultimo capitolo il modello proposto verrà applicato ad un problema reale di carattere meteorologico. Nello specifico, il modello verrà impiegato per la previsione della quantità totale di pioggia caduta nel 2021 usando le curve giornaliere della temperatura e della sua velocità di cambiamento relative all'intero anno. In Appendice, sono esposti alcuni complementi di algebra e analisi matematica, dei risultati aggiuntivi riguardanti lo studio di simulazione condotto e il codice R utilizzato per la stima del modello proposto.

Capitolo 1

I dati funzionali

L'analisi dei dati funzionali (abbreviata con l'acronimo dall'inglese FDA, per *Functional Data Analysis*) è una branca della Statistica che ha iniziato ad acquisire notevole rilevanza e popolarità soprattutto a seguito dei recenti progressi tecnologici relativi ai metodi di raccolta dei dati. Al giorno d'oggi, infatti, in numerosi contesti vengono generati dei dati su una griglia di punti tanto fitta che è più conveniente che essi vengano trattati come infinito-dimensionali piuttosto che come dati con elevata dimensionalità (Hall et al., 2006). L'analisi dei dati funzionali si occupa quindi di affrontare le nuove sfide statistiche e analitiche che conseguono a questo cambiamento di punto di vista.

In questo primo capitolo, viene fatta una breve panoramica su cosa sia un dato funzionale e come esso venga rappresentato a partire da un insieme discreto di valori misurati. Per una trattazione più approfondita a riguardo si vedano ad esempio i Capitoli 3, 4 e 5 di Ramsay & Silverman (2005).

1.1 Da dati grezzi a funzioni lisce

La variabile casuale che viene considerata come un dato funzionale non è uno scalare o un vettore ma giace su uno spazio funzionale ed è definita su infiniti punti. Un'osservazione funzionale può essere dunque definita come una funzione $x_i : \mathcal{T} \rightarrow \mathbb{R}$, il cui dominio \mathcal{T} può essere ad esempio spaziale o temporale. Nella pratica però,

si osserva per l' i -esima unità statistica un insieme discreto di n valori $y_{i,1}, \dots, y_{i,n}$, corrispondenti ad esempio ad osservazioni in una griglia di punti ravvicinati nello spazio o nel tempo, che può anche essere diversa per ciascuna unità statistica. Inoltre, i dati osservati sono frequentemente affetti da errori di misurazione o di registrazione, quindi per ogni unità, si assume per gli n dati osservati un modello con la seguente forma

$$y_j = x(t_j) + \varepsilon_j \quad j = 1, \dots, n \quad (1.1)$$

dove ε_j è il rumore e $x(t)$, con $t \in \mathcal{T}$, è la funzione liscia che si vuole stimare.

Nell'approccio classico all'analisi dei dati funzionali, la fase di stima della funzione $x(t)$ a partire dai dati osservati, viene chiamata fase di lisciamiento e consiste nella stima di un modello statistico come quello definito nell'Equazione 1.1. Per le analisi successive si assumerà che il risultato di tale modello sia il dato reale, senza tenere conto dell'errore associato alle stime. Per questo motivo, il processo di lisciamiento è particolarmente cruciale nell'analisi dei dati funzionali e deve essere tale per cui la curva stimata sia depurata dal rumore ma anche sufficientemente flessibile da cogliere i *pattern* locali presenti nei dati.

1.2 Tecniche di lisciamiento

Le tecniche di modellazione che risultano più adeguate al contesto descritto nella sezione precedente sono di tipo non parametrico e in particolare prevedono l'utilizzo di funzioni di base. Queste ultime sono uno strumento divenuto popolare grazie alla sua flessibilità e alla notevole efficienza computazionale.

Un sistema di basi M -dimensionale è un insieme di funzioni note $\{\phi_m, m = 1, \dots, M\}$, matematicamente indipendenti l'una dall'altra e che godono della proprietà di poter approssimare arbitrariamente bene qualsiasi funzione considerando una combinazione lineare di un numero sufficientemente grande di queste funzioni. Questo approccio consiste quindi nel rappresentare la funzione $x(t)$ tramite un'espansione lineare

$$x(t) = \sum_{m=1}^M \psi_m \phi_m(t) \quad (1.2)$$

in termini di M funzioni di base note. Il notevole vantaggio che deriva da una formulazione come quella in Equazione 1.2 è che è possibile esprimere funzioni non lineari e con caratteristiche specifiche locali, pur mantenendo una struttura lineare del modello rispetto alle funzioni di base.

Una volta scelto il sistema di basi, l'Equazione 1.1 può essere scritta nella forma

$$y_j = \sum_{m=1}^M \psi_m \phi_m(t_j) + \varepsilon_j \quad j = 1, \dots, n$$

e il problema di stima si riduce quindi a un classico problema di regressione lineare multipla *standard* con parametri di regressione ψ_1, \dots, ψ_m .

Quando la dimensione dell'insieme di basi M è pari al numero di punti osservati n , scegliendo opportunamente i coefficienti ψ_m , si possono interpolare esattamente i punti, cioè far sì che $y_j = x(t_j)$, per $j = 1, \dots, n$. Si comprende quindi che M è esso stesso un parametro del sistema di basi, che permette di controllare quanto la curva rappresentata sia liscia o frastagliata. Esso viene scelto in modo tale che sia grande abbastanza da catturare i *pattern* presenti in $x(t)$ e non perdere troppa informazione, ma piccolo abbastanza da regolarizzare la stima. Una volta scelta la dimensione dell'insieme di basi tenendo conto di tale compromesso, il vettore M -dimensionale di coefficienti ψ può essere stimato usando il criterio dei minimi quadrati ordinari.

Il principale limite di questo approccio di regolarizzazione basato sulla rappresentazione della funzione tramite un sistema di basi troncato, è che la complessità del modello cresce in modo discreto a seconda del valore di M . Un metodo più flessibile, che permette un controllo più fine, è basato sull'utilizzo come criterio di stima dei minimi quadrati penalizzati. Questo prevede che la stima dei coefficienti delle basi sia pari al minimo della funzione obiettivo

$$\sum_{j=1}^n \left(y_j - \sum_{m=1}^M \psi_m \phi_m(t_j) \right)^2 + \lambda \int [Lx(t)]^2 dt.$$

La penalità presente in forma additiva induce una compressione della variabilità della funzione $x(t)$ che viene controllata in modo continuo dal parametro di lisciamiento λ . Viene costruita a partire da un operatore differenziale lineare L ,

uno strumento utile a quantificare il grado di irregolarità di una curva. Per una definizione formale di tale strumento matematico si rimanda all'Appendice A.1.

1.3 Sistemi di funzioni di base

Il processo di stima della funzione liscia a partire dai dati grezzi descritto è condizionato all'insieme di funzioni di base scelto per rappresentare la funzione. Esistono diversi tipi di funzioni di base, ciascuno adatto a rappresentare funzioni con determinate caratteristiche. Tra le funzioni di base più usate vi sono le basi polinomiali, le basi Fourier, quelle indotte dalle *splines*, come ad esempio le *B-splines*, ect. In seguito sono descritti i sistemi di base che verranno utilizzati nel presente elaborato, ossia le *B-splines* e le basi di Fourier.

1.3.1 *B-splines*

Le funzioni di base di tipo *B-splines* sono funzioni polinomiali a tratti definite da ordine, *range* di variabilità e un insieme di nodi (De Boor, 1978). Le *B-splines* di ordine d sono formate da polinomi di grado $d - 1$ stimati in sub-intervalli delimitati da $M - d$ nodi interni e due nodi esterni. Il sistema di basi è quindi piuttosto complesso ma molto flessibile e per questo viene frequentemente usato per la rappresentazione dei dati funzionali non periodici.

La Figura 1.1 mostra un insieme di funzioni di base *B-splines*. Si può notare che ciascuna funzione è diversa da zero solo in un numero limitato di sub-intervalli consecutivi. Nello specifico, per un sistema di basi di ordine d , si ha che ciascuna funzione è diversa da zero in non più di d sub-intervalli consecutivi, tra loro adiacenti. Questa proprietà, detta di località o di supporto compatto, comporta un notevole guadagno computazionale. Dalla Figura 1.1, si può inoltre notare che le funzioni che sono diverse da zero negli intervalli agli estremi del *range* di variabilità in cui il sistema è definito hanno scale diverse rispetto alle funzioni centrali. Le *B-splines*, infatti, godono anche della proprietà per cui la somma delle funzioni di base in ogni punto t nel *range* di variabilità è pari a uno (Ramsay et al., 2011). Queste due proprietà di cui godono gli insiemi di basi di tipo *B-splines* risultano

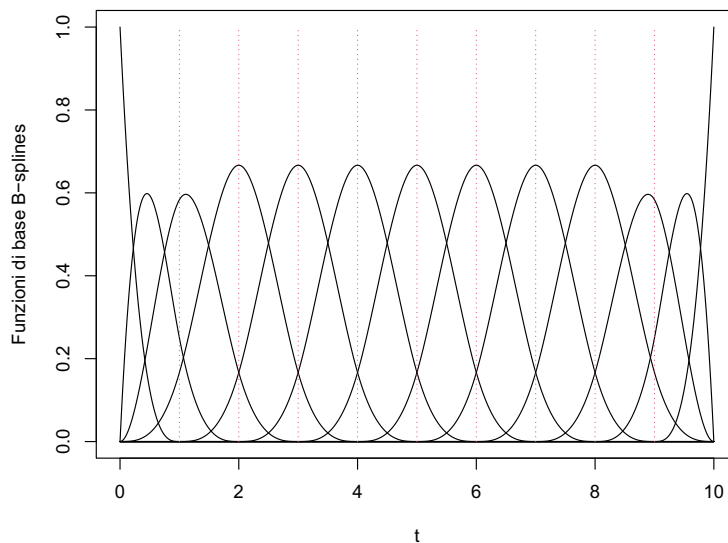


FIGURA 1.1: 13 funzioni di base di tipo *B-splines* di ordine 4 definite sull'intervallo $[0, 10]$ con 10 nodi equispaziati.

spesso particolarmente utili e fanno sì che l'utilizzo di tali sistemi sia alla base della costruzione di vari modelli statistici per l'analisi dei dati funzionali, come quelli presentati nei Capitoli 2 e 3 di questo elaborato.

Uno dei metodi più semplici e diffusi per calcolare le *B-splines* è quello basato sulla formula ricorsiva di Cox-De Boor, riportata in Appendice A.2. In R, è stata utilizzata la funzione `bs` del pacchetto `splines`, implementato da Bates & Venables (2011).

1.3.2 Basi di Fourier

Per rappresentare funzioni periodiche il sistema di basi più adeguato è quello formato dalle basi di Fourier. L'approccio consiste nel rappresentare la funzione $x(t)$ presente nell'Equazione 1.1, con una combinazione di funzioni seno e coseno adatte a cogliere le componenti sinusoidali di specifiche frequenze presenti nella funzione. Più nel dettaglio, la funzione periodica viene rappresentata mediante lo sviluppo

in serie di Fourier definito in Appendice A.3. Questo sistema di basi risulta particolarmente adeguato nel caso in cui si vogliono rappresentare delle funzioni che non hanno marcate caratteristiche locali e in cui la curvatura della funzione tende ad essere la stessa su tutto il dominio.

Per rappresentare i dati funzionali tramite basi di Fourier è stato utilizzato il pacchetto `fda` di `R`, ideato e implementato da Ramsay et al. (2011).

Capitolo 2

Modello di regressione lineare funzionale

La regressione funzionale è uno dei più utili strumenti nell'analisi dei dati funzionali e si utilizza nel contesto in cui la variabile risposta o una o più variabili esplicative possano essere considerate un dato funzionale. Il volume Horváth & Kokoszka (2012) fornisce una trattazione approfondita del tema, che negli ultimi anni è diventato sempre più centrale e di rilievo, costituendo un'area di ricerca molto attiva.

Le tecniche di regressione funzionale variano a seconda che la risposta o le covariate appartengano a spazi funzionali (e.g. funzioni o superfici) oppure euclidei (e.g. scalari o vettori). In particolare, le combinazioni per le quali occorre utilizzare approcci differenti sono le seguenti:

- (a) sia le covariate sia la risposta sono funzionali;
- (b) le covariate sono funzionali e la risposta è scalare;
- (c) le covariate sono scalari mentre la variabile risposta è funzionale.

Lo scenario (b) è quello che è stato studiato in modo più esteso nella letteratura ed è anche lo scenario d'interesse nel presente elaborato.

Il modello di regressione più semplice che si può utilizzare per analizzare dei dati di questa natura è il modello di regressione lineare funzionale. Nel presente

capitolo verrà riportata una trattazione a riguardo e in seguito verrà descritta anche una specificazione del modello, ispirata alla recente proposta di Bernardi et al. (2022), che permette di ottenere stime di facile interpretazione.

2.1 Modello di regressione funzionale classico

Il modello di regressione tradizionale con risposta scalare $Y \in \mathbb{R}$, che per semplicità si assume essere centrata in zero, e matrice delle covariate $X \in \mathbb{R}^p$, può essere espresso come

$$Y = \langle X, \beta \rangle + \varepsilon, \quad (2.1)$$

usando il prodotto interno definito nello spazio euclideo, dove β è il vettore di regressione ed ε è un vettore di errori a media nulla e con varianza finita. Sostituendo nell'Equazione 2.1, la matrice X con la funzione $X(t)$ e il vettore scalare di coefficienti di regressione con un coefficiente funzionale $\beta(t)$, si ottiene il modello di regressione funzionale (FML, acronimo inglese di *Functional Linear Model*)

$$Y = \int X(t)\beta(t)dt + \varepsilon. \quad (2.2)$$

L'integrale presente nell'Equazione 2.2 è il prodotto interno definito nello spazio funzionale tra le funzioni $X(t)$ e $\beta(t)$. Risulta dunque evidente che il modello di regressione lineare funzionale *standard* è una naturale espansione funzionale del classico modello di regressione lineare (Wang et al., 2016).

Per la stima del coefficiente di regressione funzionale $\beta(t)$ occorre utilizzare un qualche metodo di regolarizzazione al fine di ridurre il problema di collinearità e ottenere stime univocamente identificabili. A tale scopo, si possono utilizzare le tecniche di liscio descritte nella Sezione 1.2, che prevedono che la funzione da liscio, ossia $\beta(t)$, sia espressa tramite un'espansione in funzioni di base. Il coefficiente di regressione funzionale viene quindi scritto nella forma

$$\beta(t) = \sum_{m=1}^M \psi_m \phi_m(t) \quad \text{oppure} \quad \beta(t) = \Phi(t)^T \psi.$$

L'approccio della rappresentazione in funzioni di base del coefficiente funzionale impone una struttura finito-dimensionale al problema di stima, restringendo la

scelta di $\beta(t)$ a quella di un vettore di coefficienti $\psi \in \mathbb{R}^M$ associato a un prestabilito insieme di funzioni di base $\{\phi_m, m = 1, \dots, M\}$. Il modello descritto in 2.2, può infatti essere riespresso come

$$Y = Z\psi + \varepsilon$$

dove $Z = \int X(t)\Phi(t)^T dt$ e il vettore M -dimensionale di coefficienti ψ è il parametro da stimare. Se come tecnica di lisciamiento del coefficiente funzionale si sfrutta la rappresentazione mediante un sistema di basi troncato, per la stima dei parametri si può utilizzare il metodo dei minimi quadrati ordinari per il quale esiste la soluzione in forma chiusa

$$\hat{\psi} = (Z^T Z)^{-1} Z^T Y.$$

Tuttavia, per le motivazioni legate alla flessibilità del modello, descritte nella Sezione 1.2, l'approccio di stima che viene più frequentemente adoperato è quello dei minimi quadrati penalizzati. La funzione obiettivo da minimizzare per ottenere una stima del coefficiente funzionale $\beta(t)$ è quindi

$$\sum_{i=1}^N (y_i - \int x_i(t)\beta(t)dt)^2 + \lambda \int [L\beta(t)]^2 dt,$$

dove λ è il parametro di lisciamiento positivo associato alla penalizzazione che quantifica il grado di liscezza della curva. Anche la covariata funzionale viene di solito rappresentata in basi e spesso conviene utilizzare lo stesso sistema usato per il coefficiente funzionale.

2.1.1 Interpretazione del coefficiente funzionale

Come nei modelli di regressione lineare classici, anche per il modello di regressione funzionale una caratteristica fondamentale è l'interpretabilità dei coefficienti stimati. I punti del dominio in cui $\beta(t) \neq 0$ corrispondono ai punti in cui esiste una relazione tra Y e $X(t)$; al contrario, nei punti in cui $\beta(t) = 0$, non sussiste alcuna relazione. Risulta evidente che a fini interpretativi, è importante che il modello di regressione sia in grado di stimare esattamente a zero il coefficiente funzionale nelle regioni in cui non c'è un'apprezzabile relazione tra la risposta e la variabile

esplicativa.

Negli ultimi anni sono stati proposti diversi metodi di risoluzione a questo problema. Ad esempio, James et al. (2009) hanno a tale fine elaborato il modello FLiRTI (*Functional Linear Regression That's Interpretable*), basato sull'idea della selezione delle variabili applicata ad alcune derivate di $\beta(t)$. Il fatto che la penalizzazione riguardi non solo il livello del coefficiente funzionale, ma anche eventualmente alcune sue derivate, favorisce particolarmente l'interpretazione dei coefficienti funzionali risultanti perché permette di imporre vincoli di forma anche sulle regioni in cui il coefficiente è non nullo.

Recentemente, Bernardi et al. (2022) hanno introdotto il modello *LSFR* (*Locally Sparse Functional Regression*). Tale modello è applicabile al caso più generale in cui sia la variabile risposta sia la variabile esplicativa sono di tipo funzionale e sfrutta una penalizzazione di tipo *Overlap Group Lasso* sui coefficienti della rappresentazione in basi del coefficiente funzionale.

Segue nel paragrafo successivo una trattazione più approfondita del modello LSFR, adattato però al caso specifico in cui la variabile risposta è di tipo scalare e la variabile esplicativa è funzionale, ovvero al contesto di FDA studiato in questo elaborato. Nel capitolo successivo viene quindi proposta un'estensione del modello utile in presenza di due covariate funzionali con struttura di sparsità condivisa.

2.2 Modello di regressione lineare funzionale localmente sparso

Si assume per semplicità e senza perdita di generalità che ci sia un'unica covariata funzionale x_i disponibile per $i = 1, \dots, N$. La relazione tra la risposta y_i e il predittore è del tutto analoga a quella assunta dal modello di regressione funzionale classico (2.2) ed è quindi data da

$$y_i = \int x_i(t)\beta(t)dt + \varepsilon_i.$$

Il modello di regressione funzionale localmente sparso prevede però l'utilizzo di una specifica rappresentazione in basi del coefficiente funzionale $\beta(t)$ e di un criterio di stima basato su una penalizzazione che ne permetta la sparsità.

2.2.1 Rappresentazione in basi del coefficiente funzionale

Come nel modello di regressione funzionale classico, per rappresentare il coefficiente funzionale $\beta(t)$ si utilizza un'espansione in basi, che permette di esprimerlo nella forma

$$\beta(t) = \sum_{m=1}^M \psi_m \phi_m(t).$$

Si assume poi che le funzioni che compongono il sistema di basi $\{\phi_m, m = 1, \dots, M\}$ siano delle *B-splines* di ordine d . Esse, come descritto nella Sezione 1.3.1, godono della proprietà di località per cui in ciascun sub-intervallo delimitato da due nodi solo un numero limitato di funzioni di base consecutive sono diverse da zero. Ne consegue che, se in un certo sub-intervallo i coefficienti associati alle uniche funzioni di base positive dovessero essere nulli, allora anche la funzione rappresentata sarebbe nulla nell'intero spazio delimitato dai due nodi. Quindi, se nel vettore di coefficienti ψ delle basi di tipo *B-splines* è presente un *pattern* di sparsità adeguato, allora verrà indotta una struttura di sparsità anche nel corrispondente oggetto funzionale rappresentato in basi, ossia $\beta(t)$. In particolare, siano $\tau_1 < \dots < \tau_m < \dots < \tau_{M-d+2}$, con $\tau_m \in \mathcal{T}$, i nodi che definiscono i sub-intervalli $\mathcal{T}_m = (\tau_m, \tau_{m+1})$. Allora, per ottenere $\beta(t) = 0$ per ciascun $t \in \mathcal{T}_m$, è sufficiente che siano pari a zero tutti i coefficienti $\psi_{m'}$, con $m' = m, \dots, m + d - 1$, ossia che il blocco di elementi di ψ di lunghezza d associato alla regione \mathcal{T}_m , sia interamente pari a zero. Da ciò si evince che per ottenere stime sparse del coefficiente $\beta(t)$ è utile dividere il vettore ψ dei coefficienti delle *B-splines* in blocchi di lunghezza d e utilizzare una penalizzazione che tenga conto di tale ripartizione. A tale fine si può sfruttare una penalizzazione di tipo *Overlap group Lasso*, discussa nella seguente sezione.

2.2.2 Penalizzazione che induce sparsità

L'*Overlap Group Lasso* è una penalizzazione che induce una selezione delle variabili strutturata, inglobando nel modello conoscenza a priori relativa ai *pattern* di sparsità attesi. Jenatton et al. (2011) hanno proposto un metodo per costruire penalizzazioni di questo tipo, che inserite nella funzione obiettivo di un classico modello di regressione lineare, conducono a specifici modelli lineari sparsi, in cui la previsione dell'*output* è ottenuta combinando linearmente solo specifici sottoinsiemi delle variabili che descrivono i dati. La penalizzazione proposta è costruita sommando norme relative a gruppi sovrapposti di variabili, che vengono definiti a priori in modo da indurre la sparsità attesa. Quindi, per utilizzare la penalizzazione di tipo *Overlap group lasso* per la stima dei coefficienti del modello di regressione lineare funzionale occorre innanzitutto specificare un'adeguata struttura a gruppi per le funzioni di base utilizzate nella rappresentazione del coefficiente funzionale.

Si definisce dunque una sequenza di blocchi parzialmente sovrapposti di dimensione d , pari all'ordine delle *B-splines* utilizzate, e formati dagli elementi del vettore di coefficienti ψ associati alle basi. Sia $b = 1, \dots, B$ un indicatore di blocco, con $B = M - d + 1$ che denota il numero totale di blocchi. Il generico b -esimo insieme contiene i coefficienti che appartengono all'insieme

$$\left\{ \psi_m : m = m^*, \dots, m^* + d - 1, m^* = \left\lceil \frac{b}{M - d + 1} \right\rceil \right\}.$$

Il primo gruppo quindi è formato dai primi d elementi di ψ , mentre i successivi si possono identificare scorrendo gli altri elementi di ψ tramite *steps* di lunghezza uno. Un esempio di una tale suddivisione degli elementi di ψ in gruppi parzialmente sovrapposti è visibile nella Figura 2.1.

Una penalizzazione che tiene conto della suddivisione in gruppi proposta, favorisce effettivamente una struttura di sparsità in $\beta(t)$ che rende il modello più facilmente interpretabile. Nello specifico, sia \mathcal{T}_0 l'insieme dei punti del dominio in cui il coefficiente funzionale viene stimato esattamente pari a zero: $\mathcal{T}_0 = \{t \in \mathcal{T} | \beta(t) = 0\}$. La penalizzazione, favorisce che l'insieme \mathcal{T}_0 sia formato dall'unione di intervalli \mathcal{T}_m e ciò comporta che $\beta(t)$ sia nullo in punti vicini del dominio. A titolo di esempio si osservi la Figura 2.2. Nel primo pannello vengono rappresen-

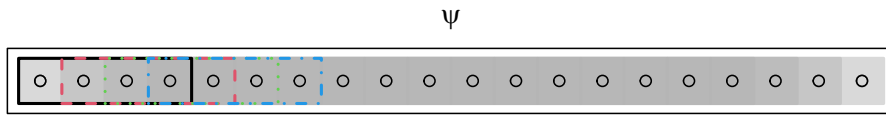


FIGURA 2.1: Suddivisione in blocchi del coefficiente ψ ($M = 20$), associato a un sistema di *B-splines* di ordine 4. Le zone centrali appaiono più scure perché comprendono coefficienti che appartengono a un maggior numero di gruppi. I rettangoli mostrano i primi 4 gruppi.

tati gli elementi di ψ , il vettore dei coefficienti associato a un sistema di basi di dimensione $M = 20$. I rettangoli colorati indicano i gruppi sovrapposti di coefficienti stimati esattamente a zero. Nel pannello sottostante, si osserva l'insieme \mathcal{T} suddiviso negli intervalli identificati dai 16 nodi interni. La struttura di sparsità dei coefficienti delle basi ψ induce sparsità nel coefficiente funzionale $\beta(t)$ e la regione \mathcal{T}_0 dove $\beta(t)$ viene stimato esattamente a zero, è quella colorata in grigio. Si può notare che è effettivamente scrivibile come l'unione di regioni \mathcal{T}_m consecutive: $\mathcal{T}_0 = \mathcal{T}_5 \cup \mathcal{T}_6 \cup \mathcal{T}_7 \cup \mathcal{T}_8$.

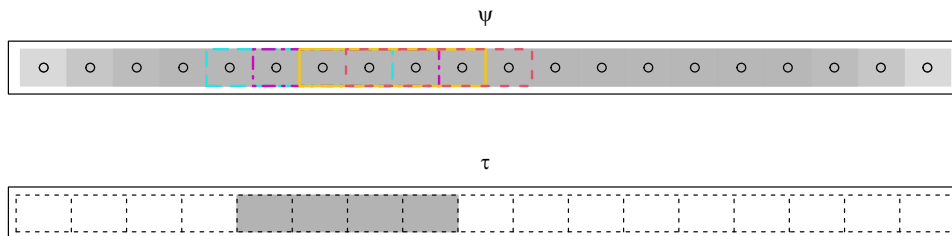


FIGURA 2.2: Nel primo pannello, il vettore di coefficienti ψ , con i gruppi stimati pari a 0 contrassegnati dai rettangoli colorati. Nel secondo pannello, l'insieme \mathcal{T} suddiviso in intervalli e la regione \mathcal{T}_0 in grigio che evidenzia la corrispondente struttura di sparsità indotta sul coefficiente $\beta(t)$.

2.2.3 Funzione obiettivo

Si definisce la funzione obiettivo da minimizzare per ottenere la stima di ψ come

$$\frac{1}{2} \sum_{i=1}^n \left(y_i - \int x_i(t) \beta(t) dt \right)^2 + \lambda \sum_{b=1}^{B+1} \|c_b \odot \psi\|_2 \quad (2.3)$$

dove la prima sommatoria è un indice di bontà di adattamento mentre la seconda è la penalizzazione adeguata a favorire la sparsità di $\beta(t)$, che viene regolata tramite il parametro di penalizzazione positivo λ . La funzione di penalizzazione è costituita dalla somma delle $B+1$ norme euclidee $\|c_b \odot \psi\|_2$, dove \odot è il prodotto di Hadamard definito in Appendice A.4. Ciascuna norma euclidea è riferita a un gruppo di coefficienti che vengono penalizzati insieme e i primi B gruppi costituiscono la suddivisione in blocchi parzialmente sovrapposti descritta sopra, mentre l'ultimo è complessivo di tutti gli elementi di ψ e viene inserito nella struttura a gruppi perché permette di migliorare le proprietà teoriche associate al modello di regressione. Il vettore c_b di lunghezza M riferito al b -esimo gruppo contiene dei pesi che vengono calcolati in modo deterministico e sono utili a bilanciare il fatto che gli elementi agli estremi di ψ compaiono in meno gruppi rispetto a quelli centrali, come si può notare osservando la Figura 2.1. In particolare, il generico vettore c_b è definito come $c_b = s_b \odot C$, dove s_b è il vettore di selezione lungo M il cui generico elemento di posizione m vale 1 se il corrispondente elemento in posizione m del vettore di parametri ψ appartiene al b -esimo gruppo e 0 altrimenti. Il generico elemento c_m del vettore C , invece, è pari a

$$c_m = \left(\sum_{b=1}^{B+1} s_m^{(b)} \right)^{-1}.$$

Si può dimostrare che il modello di regressione funzionale localmente sparso definito, che sfrutta alcune proprietà delle B -splines e la penalizzazione *Overlap Group Lasso* costruita *ad hoc*, gode di interessanti proprietà, relative in particolare alla coesistenza delle stime. Nello specifico, sotto l'ipotesi di corretta specificazione del modello e di incorrelazione e normalità degli errori ε_i con varianza $\sigma^2 > 0$, il metodo riesce ad identificare correttamente le regioni dove il coefficiente funzionale

è realmente nullo. Per la dimostrazione formale e ulteriori approfondimenti a riguardo, si rimanda a (Bernardi et al., 2022, Teoremi 2 e 3).

Capitolo 3

Modello di regressione lineare funzionale con struttura di sparsità condivisa

Nel presente capitolo verrà proposto un particolare modello di regressione funzionale, che costituisce un'estensione del modello di regressione funzionale localmente sparso descritto nel precedente capitolo. Si tratta di un modello utile quando vi sono due variabili esplicative funzionali definite sullo stesso dominio e in alcune regioni condivise di tale dominio è ragionevole ritenere che non sussista alcuna relazione con la risposta. Nel capitolo verranno innanzitutto descritte le proprietà auspicabili per un modello da applicare in tale contesto e in seguito verrà specificato il modello proposto.

3.1 Proprietà desiderate

Si consideri il caso in cui si voglia prevedere una variabile risposta scalare Y usando due covariate funzionali $X_1(t)$, con $t \in \mathcal{T}$, e $X_2(t)$, con stesso dominio \mathcal{T} . Si assuma, inoltre, che vi siano delle regioni del dominio, $\mathcal{T}_{1,0}$ e $\mathcal{T}_{2,0}$, in cui la variabile risposta non è in relazione con le rispettive due variabili esplicative. Un problema previsivo di questo tipo può essere riscontrato in diversi ambiti applicativi, come

ad esempio quelli meteorologico, economico, sportivo e sociale. Inoltre, capita talvolta che le due variabili esplicative condividano la stessa struttura di sparsità e quindi che le regioni in cui i due coefficienti funzionali sono nulli, rispettivamente $\mathcal{T}_{1,0}$ e $\mathcal{T}_{2,0}$, siano quasi coincidenti.

Si pensi ad esempio a dei dati relativi a una competizione sportiva, come una gara di ciclismo che si svolge in un percorso in cui vi sono dei tratti percorribili senza difficoltà e altri invece molto più insidiosi. Si immagini di avere a disposizione due covariate funzionali rilevate lungo il corso della gara, come ad esempio la cadenza delle pedalate e la frequenza cardiaca dei ciclisti, e che si vogliano utilizzare questi predittori al fine di stimare il tempo di arrivo di ciascun concorrente. È ragionevole pensare che entrambe le covariate siano utili nella previsione della risposta solo e soprattutto nei tratti più decisivi del percorso.

Un altro esempio può essere riferito a un problema in ambito meteorologico, in cui si vuole studiare la relazione tra gli andamenti della temperatura e della pressione atmosferica al suolo nel corso di un intero anno e il numero totale di eventi temporaleschi che avvengono nello stesso anno di osservazione in una certa zona climatica. Dal momento che il temporale è un fenomeno che si verifica per lo più nei mesi estivi, in un modello di regressione funzionale per la previsione del numero di rovesci temporaleschi potrebbe esistere una struttura di sparsità condivisa nei coefficienti associati alle due covariate funzionali relative a temperatura e pressione.

In questi casi è auspicabile che il modello di regressione utilizzato riesca a individuare le regioni in cui non esiste relazione tra la risposta e le covariate, in modo che i coefficienti funzionali, e quindi lo stesso fenomeno spiegato dal modello, siano più facilmente interpretabili. Si desidera quindi definire e stimare un modello di regressione per dati funzionali che, oltre a mantenere buone capacità previsive, favorisca un *pattern* di sparsità condiviso dai coefficienti funzionali associati alle due covariate.

Un modello di regressione con queste caratteristiche può risultare utile anche nel caso in cui si abbia a disposizione un'unica covariata funzionale ma si voglia utilizzare per la previsione della risposta anche una sua derivata. Può infatti capitare che se il livello della curva è in relazione con la variabile risposta solo in

alcune regioni del dominio, anche la derivata della curva sia rilevante solo nelle suddette regioni. Per questo, un modello che favorisce stessa struttura di sparsità nelle covariate (pur non imponendola) potrebbe essere utile per l'interpretazione del fenomeno. Ad esempio, nel Capitolo 5, per prevedere il numero totale di millimetri di pioggia caduti nel 2021 in alcune centraline disposte in Veneto, si utilizzano la curva giornaliera della temperatura dello stesso anno e quella della velocità con cui cambia la temperatura, ossia la sua derivata prima.

3.2 Specificazione del modello

Il modello di regressione lineare funzionale adatto al caso descritto nel paragrafo precedente assume che la relazione tra la variabile risposta e le due covariate funzionali sia della forma

$$y_i = \int x_{i1}(t)\beta_1(t)dt + \int x_{i2}(t)\beta_2(t)dt + \varepsilon_i, \quad (3.1)$$

con $i = 1, \dots, N$. Analogamente a quanto effettuato per il modello di regressione lineare sparso descritto nel Paragrafo 2.2, i coefficienti funzionali vengono rappresentati tramite un insieme di funzioni di base di tipo *B-splines* di ordine d . I coefficienti possono dunque essere scritti come

$$\beta_1(t) = \sum_{m=1}^M \psi_{1,m} \phi_m(t) = \phi(t)^T \psi_1 \quad \text{e} \quad \beta_2(t) = \sum_{m=1}^M \psi_{2,m} \phi_m(t) = \phi(t)^T \psi_2.$$

Si noti che viene utilizzato lo stesso sistema di basi M -dimensionale per entrambi i coefficienti funzionali.

La funzione obiettivo da minimizzare per ottenere la stima di $\psi = (\psi_1, \psi_2)$ è formata da due componenti additive:

$$\frac{1}{2} \sum_{i=1}^n (y_i - \int x_{i1}(t)\beta_1(t)dt - \int x_{i2}(t)\beta_2(t)dt)^2 + \lambda \Omega(\psi). \quad (3.2)$$

Come di consueto, la prima componente è una misura dell'errore di previsione mentre la seconda è una funzione di penalizzazione. Come descritto nel paragrafo precedente, quest'ultima deve essere costruita in modo tale da consentire che:

- la stima dei due coefficienti funzionali sia sparsa;
- il *pattern* di sparsità sia condiviso tra i due coefficienti.

La prima delle due proprietà desiderate è ottenibile usando una penalizzazione analoga a quella introdotta per il modello di regressione localmente sparso, che sfrutta la proprietà di località delle *B-splines* con cui sono espressi i coefficienti funzionali. Per far sì che venga selezionato un giusto *pattern* di coefficienti delle basi in modo da permettere la struttura di sparsità condivisa, occorre inoltre che la funzione di penalizzazione presente nell'Equazione 3.2 sia ancora di tipo *Overlap Group Lasso*. Occorre però definire un'adeguata suddivisione in gruppi degli elementi dei due vettori ψ_1 e ψ_2 associati alla rappresentazione in basi dei due coefficienti funzionali.

3.2.1 Penalizzazione che induce sparsità con *pattern* condiviso

Si specifica dunque la struttura a gruppi parzialmente sovrapposti che permette di costruire la penalizzazione *Overlap Group Lasso* adeguata a ottenere uno stimatore con le due caratteristiche desiderate.

I primi $B = M - d + 1$ gruppi contengono solo elementi di ψ_1 , ossia del vettore di coefficienti delle basi relativo al primo coefficiente funzionale, nonché i primi M elementi di $\psi = (\psi_1, \psi_2)$. I B gruppi parzialmente sovrapposti vengono definiti in modo conforme a quanto descritto per il caso con una sola covariata funzionale. Il generico b -esimo blocco, con $b = 1, \dots, B$, infatti, contiene i coefficienti che appartengono all'insieme

$$\left\{ \psi_{1m} : m = m^*, \dots, m^* + d - 1, m^* = \left\lceil \frac{b}{M - d + 1} \right\rceil \right\}.$$

Questi primi B gruppi, quindi, permettono la sparsità del primo coefficiente funzionale.

Analogamente vengono definiti altri B blocchi, riferiti però al secondo coefficiente funzionale, quindi agli elementi di ψ_2 . Pertanto, il generico b -esimo blocco, con $b = B + 1, \dots, 2B$, contiene i coefficienti che appartengono all'insieme

$$\left\{ \psi_{2m} : m = m^*, \dots, m^* + d - 1, m^* = \left\lceil \frac{b}{M - d + 1} \right\rceil \right\}.$$

Una penalizzazione *Overlap group Lasso* che tenga conto di questi primi $2B$ gruppi definiti, consentirebbe stime sparse di entrambi i coefficienti funzionali, senza però in alcun modo favorire che il *pattern* di sparsità sia il medesimo tra i due coefficienti. A tale fine, occorre tenere conto nella penalizzazione di altri B gruppi. Il generico b -esimo blocco, con $b = 2B + 1, \dots, 3B$, contiene i coefficienti che appartengono all'insieme

$$\left\{ \psi_{jm} : m = m^*, \dots, m^* + d - 1, m^* = \left\lceil \frac{b}{M - d + 1} \right\rceil, j = 1, 2 \right\}.$$

Questi nuovi B gruppi sono dunque formati ciascuno da d elementi di ψ_1 e d elementi di ψ_2 , posti nella stessa posizione nei due vettori. Quindi, i d elementi di ψ_1 e ψ_2 appartenenti allo stesso gruppo sono riferiti alle stesse funzioni di base di tipo *B-splines* con cui i due coefficienti sono rappresentati. Pertanto la nullità dell'intero gruppo di coefficienti comporta la nullità dei coefficienti funzioni $\beta_1(t)$ e $\beta_2(t)$ nello stesso intervallo del loro supporto \mathcal{T} .

Infine, si inseriscono nella struttura a blocchi anche due ultimi gruppi: uno contenente tutti gli elementi di ψ_1 e uno con tutti gli elementi di ψ_2 . Il numero totale di gruppi definiti è quindi pari a $3B + 2$, con $B = M - d + 1$. La Figura 3.1 mostra una rappresentazione grafica utile a visualizzare come i gruppi di elementi di ψ sono stati definiti.

Si definisce quindi la funzione obiettivo da minimizzare per ottenere la stima di ψ come

$$\frac{1}{2} \sum_{i=1}^n \left(y_i - \int x_{i1}(t) \beta_1(t) dt - \int x_{i2}(t) \beta_2(t) dt \right)^2 + \lambda \sum_{b=1}^{3B+2} \|c_b \odot \psi\|_2. \quad (3.3)$$

Si noti che la funzione obiettivo ha la stessa struttura di quella espressa nell'Equazione 2.3, che si utilizza nel modello di regressione localmente sparso con una sola covariata funzionale. In questo caso però il vettore ψ è di lunghezza non più pari a M , ma pari a $2M$ perché contiene i coefficienti delle basi di entrambi i coefficienti funzionali. Anche i vettori c_b sono quindi di lunghezza $2M$ ma sono costruiti in

modo analogo a quanto indicato nella Sezione 2.2.3 per il modello con una sola covariata: sono pari al prodotto di Hadamard tra un vettore di selezione s_b e un vettore di pesi C . Il vettore di pesi C viene calcolato in modo deterministico e ha la funzione di compensare il fatto che alcuni elementi di ψ compaiono in un numero minore di gruppi rispetto ad altri.

3.3 Controparte empirica delle quantità descritte

Si assume di osservare, per un campione di numerosità N , il valore y_i della variabile risposta scalare, con $i = 1, \dots, N$. Si osservano anche le due covariate funzionali $x_{i1}(t)$ e $x_{i2}(t)$ in due griglie di punti potenzialmente diverse. Si assume per semplicità ma senza perdita di generalità che queste due griglie di punti siano della medesima lunghezza n . Si definiscono allora le matrici X_1 e X_2 contenenti i punti osservati delle covariate in corrispondenza di ciascuna unità statistica e di ciascun punto della griglia. Sia poi Φ la matrice $M \times n$ definita come

$$\Phi = \begin{pmatrix} \phi_1(s_1) & \cdots & \phi_1(s_n) \\ \vdots & & \vdots \\ \phi_m(s_1) & \cdots & \phi_m(s_n) \\ \vdots & & \vdots \\ \phi_M(s_1) & \cdots & \phi_M(s_n) \end{pmatrix}$$

dove s_1, \dots, s_n sono i punti in cui sono osservate le funzioni. Indicando con Y il vettore di osservazioni della variabile risposta e con ε il vettore di errori, il modello di regressione lineare funzionale descritto dall'Equazione 3.1 può essere allora scritto in forma matriciale come

$$Y = X_1 \Phi^T \psi_1 + X_2 \Phi^T \psi_2 + \varepsilon = Z \psi + \varepsilon,$$

dove $Z = (X_1 \Phi^T, X_2 \Phi^T)$ è la matrice del disegno e $\psi = (\psi_1, \psi_2)$ è il vettore dei parametri da stimare. Il problema di ottimizzazione 3.3 diventa

$$\hat{\psi}_\lambda = \arg \min_{\psi} l(\psi) \quad l(\psi) = \frac{1}{2} \|Y - Z\psi\|_2^2 + \lambda \sum_{b=1}^{3B+2} \|D_b \psi\|_2, \quad (3.4)$$

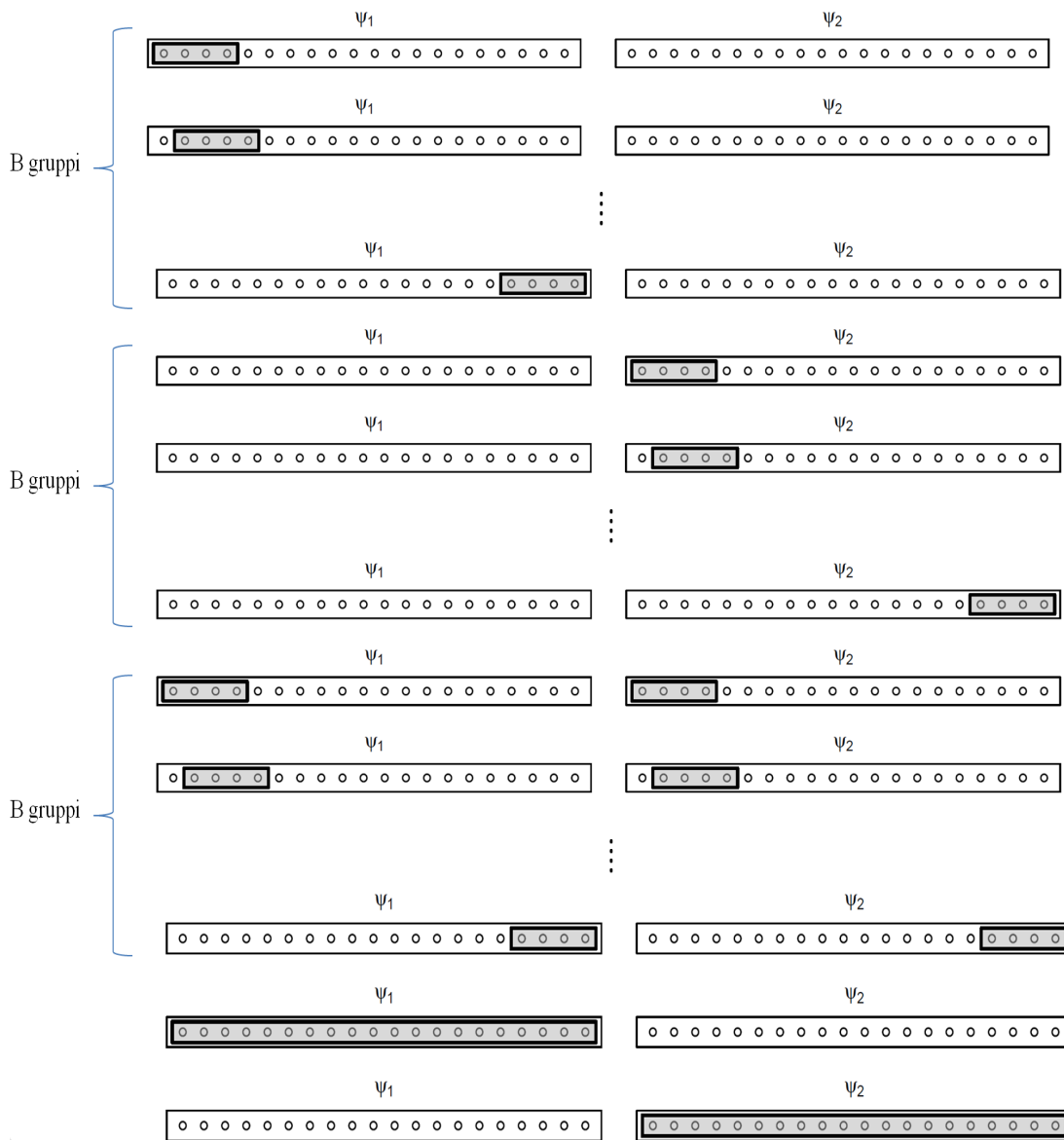


FIGURA 3.1: In ciascuna riga della Figura è rappresentato uno dei $3B + 2$ gruppi definiti per la penalizzazione *Overlap Group Lasso*. Gli elementi di ψ_1 e ψ_2 che appartengono al gruppo sono racchiusi in un rettangolo nero.

dove $D_b = \text{diag}(c_b)$ è una matrice diagonale di dimensione $2M \times 2M$ che ha nella diagonale principale gli elementi del vettore c_b , definito nella sezione precedente.

Il problema di ottimizzazione non ha soluzione esplicita e l'algoritmo *standard* di discesa del gradiente non risulta adeguato. A causa della non separabilità della penalizzazione *Overlap Group Lasso*, che non permette l'aggiornamento individuale dei parametri, l'algoritmo corre infatti il rischio di bloccarsi frequentemente in minimi locali (Jacob et al., 2009). Si ricorre allora ad algoritmi di ottimizzazione numerica. In particolare, per l'implementazione del modello è stata utilizzata la funzione `nlminb` di R.

3.4 Trasformazione delle covariate funzionali

I metodi di *shrinkage* basati su penalizzazioni come quello proposto, non sono invarianti rispetto alla scala con cui sono misurate le covariate. Per il modello con due covariate funzionali con struttura di sparsità condivisa il problema risulta essere notevolmente rilevante. Infatti, essendo presenti due diverse variabili esplicative, se queste sono state rilevate su due scale diverse occorre applicare una trasformazione che le renda confrontabili in termini numerici, in modo che la penalizzazione relativa ai due vettori di coefficienti delle basi sia equa.

Nei modelli di regressione penalizzati *standard*, come la *Ridge* o il *Lasso*, prima di calcolare la soluzione del problema di stima, viene standardizzata la matrice del disegno (Hastie et al., 2009). Tuttavia, nel caso in esame, non è ragionevole applicare tale trasformazione. La matrice del disegno Z , infatti, è costruita moltiplicando tra loro le matrici con le osservazioni relative alle covariate e la matrice con i valori delle funzioni di base calcolate negli stessi punti di osservazione. Standardizzando la matrice del disegno Z si perderebbero le utili proprietà delle *B-splines* descritte nella Sezione 1.3.1, che stanno alla base della costruzione del modello.

Una possibile soluzione potrebbe dunque essere quella di standardizzare le due covariate funzionali $X_1(t)$ e $X_2(t)$, anziché la matrice del disegno Z . Standardizzare un insieme di curve osservate $x_i(t)$, $i = 1, \dots, N$ significa applicare la seguente

trasformazione:

$$\tilde{x}_i(t) = \frac{x_i(t) - \bar{x}(t)}{v(i, i)},$$

dove $\bar{x}(t) = n^{-1} \sum_{i=1}^n x_i(t)$ è la funzione media campionaria, mentre $v(i, i) = (n-1)^{-1} \sum_{i=1}^n (x_i(t) - \bar{x}(t))^2$ è la funzione di varianza campionaria. In questo modo l'insieme di curve ottenute sarà centrato sullo zero e applicando la trasformazione a entrambe le covariate si ottengono curve tra loro confrontabili. Così facendo si perde però l'informazione relativa all'andamento medio delle due curve, fondamentale per interpretare il fenomeno oggetto di studio.

Si è deciso allora di applicare alle due covariate una diversa trasformazione, che faccia sì che le funzioni osservate per le due covariate condividano lo stesso *range* di variabilità e mantengano comunque il loro andamento medio. Una trasformazione che permette di ottenere curve con queste caratteristiche è la normalizzazione, detta anche *Min-Max scaling*, che riduce il *range* originale delle curve al solo intervallo $[0, 1]$ (Han et al., 2022, Paragrafo 2.5.1). Normalizzare un insieme di curve osservate $x_i(t)$, $i = 1, \dots, N$ significa applicare la seguente trasformazione:

$$\tilde{x}_i(t) = \frac{x_i(t) - \min_{i=1 \dots N, t \in \mathcal{T}} x_i(t)}{\max_{i=1 \dots N, t \in \mathcal{T}} x_i(t) - \min_{i=1 \dots N, t \in \mathcal{T}} x_i(t)}.$$

Dopo aver normalizzato entrambe le covariate si risolve il problema di ottimizzazione 3.4, ottenendo una stima del vettore ψ , da cui è possibile ricavare $\hat{\beta}_1(t)$ e $\hat{\beta}_2(t)$. Per poter agevolmente interpretare le stime dei coefficienti funzionali, si applica infine la trasformazione inversa

$$\tilde{\beta}_j(t) = \frac{\hat{\beta}_j(t)}{\max_{i=1 \dots N, t \in \mathcal{T}} x_{ji} - \min_{i=1 \dots N, t \in \mathcal{T}} x_{ji}(t)} \quad j = 1, 2 \quad (3.5)$$

che riporta nella scala originale i coefficienti stimati con le variabili standardizzate.

Capitolo 4

Studio di simulazione

Viene ora condotto uno studio di simulazione per valutare le *performance* empiriche del modello e metterne in luce i punti di forza e debolezza.

In questo capitolo verranno innanzitutto descritte la struttura dello studio e le metriche con cui verrà misurata l'accuratezza delle stime. Saranno poi riportati i risultati empirici ottenuti sia con il modello proposto sia con altri modelli già esistenti, al fine di fare un confronto che permetta di cogliere i vantaggi e gli svantaggi del modello stesso.

4.1 Parametri dello studio

Le due variabili esplicative funzionali $X_1(t)$ e $X_2(t)$ vengono costruite usando lo stesso sistema di basi di tipo *B-splines* di ordine 4, con 15 gradi di libertà e nodi equispaziati. Per entrambe le covariate, i coefficienti associati a ciascuna base sono generati casualmente ad ogni simulazione. Per semplicità, si assume inoltre che tutte le curve vengano osservate su una griglia equispaziata di 100 punti.

Sono stati considerati due diversi *pattern* di sparsità dei coefficienti di regressione funzionale: il primo in cui la struttura di sparsità è condivisa dai due coefficienti e il secondo in cui invece le zone di nullità delle due curve sono diverse. È d'interesse, infatti, valutare il modello sia nel caso specifico per il quale esso è stato ideato, sia in contesti più generali in cui vi sono sì delle zone nulle in entram-

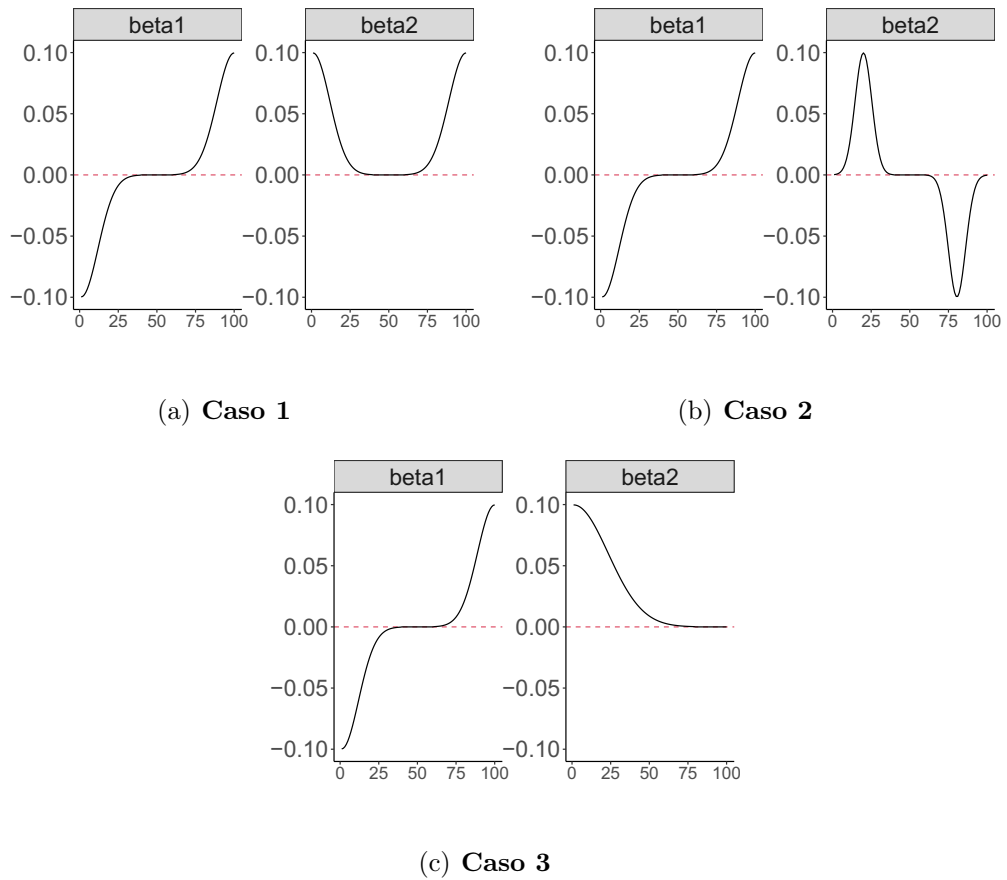


FIGURA 4.1: Tre diverse forme dei coefficienti funzionali $\beta_1(t)$ e $\beta_2(t)$ usati per la generazione dei dati nello studio di simulazione.

bi i coefficienti funzionali, ma queste non necessariamente coincidono. A questo scopo sono state considerate tre diverse forme dei coefficienti funzionali, che sono rappresentate nella Figura 4.1:

- **Caso 1:** entrambi i coefficienti funzionali sono nulli nella regione centrale del loro dominio e man mano che ci si allontana da tale regione, i coefficienti progressivamente si discostano dallo zero;
- **Caso 2:** entrambi i coefficienti funzionali sono nulli nella regione centrale del loro dominio e man mano che ci si allontana da tale regione, il primo coefficiente progressivamente si discosta dallo zero. Il secondo, invece, presenta

delle regioni prossime allo zero anche agli estremi del dominio;

- **Caso 3:** i due coefficienti funzionali sono pari a zero in due diverse regioni del dominio.

Come richiesto nella specificazione del modello, entrambi i coefficienti funzionali $\beta_1(t)$ e $\beta_2(t)$ vengono rappresentati con lo stesso insieme di funzioni di base di tipo *B-splines*. Nello specifico, si utilizza un sistema con $M = 20$ funzioni di base di ordine $d = 4$. Il vettore di coefficienti $\psi = (\psi_1, \psi_2)$ associati alla rappresentazione in basi di $\beta_1(t)$ e $\beta_2(t)$ è quindi di lunghezza $2M = 40$ ed è il vettore dei parametri che devono essere stimati.

Condizionatamente all'andamento del coefficiente funzionale considerato e alle variabili esplicative generate, viene costruita la variabile risposta sfruttando la relazione

$$y_i = \sum_{j=1}^2 \int x_{ij}(t) \beta_j(t) dt + e_i \quad i = 1, \dots, N.$$

dove il vettore di errori (e_1, \dots, e_N) , viene generato casualmente a ogni simulazione dalla distribuzione normale di media nulla e varianza tale per cui il rapporto *signal-to-noise* sia pari a un valore prestabilito.

Vengono considerati per ciascuna delle tre forme dei coefficienti di regressione descritte, tre diversi scenari di simulazione a seconda della numerosità campionaria e del rapporto segnale-rumore (si veda la Tabella 4.1).

TABELLA 4.1: Tre diversi scenari di simulazione, caratterizzati dai valori della numerosità campionaria N e dal rapporto segnale-rumore SNR .

Scenario	N	SNR
S1	50	4
S2	50	2
S3	150	4

4.2 Modelli alternativi

I risultati ottenuti dal modello proposto, vengono confrontati con quelli ottenuti tramite altri modelli di regressione funzionale regolarizzati. In particolare, per ciascuna combinazione dei parametri dello studio, la stima dei coefficienti ψ_1 e ψ_2 della rappresentazione in basi viene ottenuta minimizzando la funzione obiettivo 3.2, dove però la componente additiva relativa alla penalizzazione, $\lambda\Omega(\psi)$, viene sostituita con funzioni più semplici. In particolare, si considerano funzioni di penalizzazione di tipo *Lasso*, *Ridge* ed *Elastic-net* (Hastie et al., 2015). La stima del modello con queste penalità, viene ottenuta in **R** tramite il pacchetto `glmnet`, implementato da Friedman et al. (2010). Infine, è stato stimato anche un modello con una penalizzazione che permette una sparsità strutturata per entrambi i coefficienti funzionali, ma che, a differenza di quella del modello proposto, non favorisce che tale struttura sia la medesima per le due covariate. Si tratta del modello di regressione localmente sparso introdotto da Bernardi et al. (2022) e adattato al caso con due covariate funzionali, che differisce dal modello che permette struttura di sparsità condivisa per l'assenza degli ultimi B gruppi composti da elementi di entrambi i vettori di coefficienti.

È ragionevole aspettarsi che le stime dei coefficienti funzionali ottenute tramite la minimizzazione *Ridge* siano accurate nelle regioni in cui tali coefficienti sono diversi da zero. Questo metodo di stima, però, non permette la stima sparsa dei coefficienti funzionali e per questo ci si aspettano scarse prestazioni nelle zone in cui i coefficienti sono nulli. Il metodo di stima *Lasso*, invece, permette stime nulle dei coefficienti scalari associati alle funzioni di base, quindi è possibile che in alcune regioni i coefficienti funzionali siano stimati esattamente pari a zero, anche se, come sottolineato nel Paragrafo 2.2.1, non tutti i *pattern* di sparsità del vettore dei coefficienti delle basi implicano sparsità anche nel coefficiente funzionale corrispondente. Ad ogni modo, fatta eccezione per gli approcci che consentono sparsità strutturata, il modello che utilizza la stima *Lasso* è quello che ci si aspetta riesca a cogliere meglio le zone nulle dei coefficienti funzionali. Le stime ottenute tramite il modello con penalizzazione *Elastic-net*, invece, avranno caratteristiche intermedie rispetto a quelle ottenute con i due modelli appena citati. Infine, il

modello di regressione localmente sparso si immagina fornisca stime molto simili a quelle ottenute con la sua estensione proposta in questo elaborato. Nel caso in cui la struttura di sparsità sia realmente condivisa nelle due covariate, però, ci si aspetta che il modello che riesce a identificare meglio il corretto *pattern* di sparsità sia il secondo.

Tutti i metodi di stima considerati presentano nella funzione obiettivo da minimizzare un parametro di regolazione $\lambda > 0$, che controlla la quantità di *shrinkage*. All'aumentare di λ , infatti, i coefficienti ψ vengono sempre più ridotti verso lo zero (Hastie et al., 2009). La scelta del valore del parametro di regolazione deve tener conto del compromesso tra varianza e distorsione, in modo da evitare il sovraadattamento del modello ai dati ma allo stesso tempo ottenere stime non troppo distorte (Azzalini & Scarpa, 2009). A tale fine, si stima il modello su un campione di dati generati di dimensione N per una griglia fine di λ e si calcola l'errore di previsione su un secondo insieme di dati simulati, formato da 200 unità statistiche. La stima finale del parametro, $\hat{\psi}$, è dunque quella corrispondente al modello con il valore di λ che minimizza l'errore di previsione nell'insieme di verifica:

$$\hat{\psi}_\lambda = (\hat{\psi}_{1,\lambda}, \hat{\psi}_{2,\lambda}) = \arg \min_\lambda \sum_{i=1}^{200} \left(y_i - \sum_{j=1}^2 \int x_{ij}(t) \Phi(t)^T \psi_j dt \right)^2.$$

Per il modello *Elastic Net*, inoltre, viene scelto con questo approccio anche il parametro di *mixing* α relativo alle proporzioni di penalizzazione l_1 ed l_2 .

Infine, per confrontare l'accuratezza delle stime ottenute con diversi metodi e in diversi scenari di simulazione è necessario scegliere delle metriche di valutazione adeguate. In particolare, si vogliono valutare le *performance* dei modelli sia nelle regioni in cui i coefficienti funzionali sono realmente nulli sia in quelle in cui essi hanno un andamento che si discosta dallo zero. Per questo vengono usate due diverse metriche, ciascuna delle quali viene calcolata separatamente per $\beta_1(t)$ e $\beta_2(t)$. Visto che i coefficienti di regressione sono delle funzioni, si usano due particolari misure di errore quadratico integrato parziale: una calcolata nella regione in cui il coefficiente funzionale è realmente nullo e una nella regione in cui è diverso da zero. La prima misura quanto il modello è riuscito a catturare il corretto *pattern* di sparsità mentre la seconda quantifica l'errore di stima nelle regioni in cui il

coefficiente è una curva liscia. Quindi, per ciascuno dei due coefficienti funzionali ($j = 1, 2$), le metriche di valutazione saranno

$$ISE_0(\hat{\beta}_j(t)) = \int_{\mathcal{T}_{0,j}} \left(\beta_j(t) - \hat{\beta}_j(t) \right)^2 dt,$$

$$ISE_1(\hat{\beta}_j(t)) = \int_{\mathcal{T}_{0,j}^c} \left(\beta_j(t) - \hat{\beta}_j(t) \right)^2 dt.$$

Inoltre, si considera anche una misura di errore globale, pari a una combinazione lineare delle metriche ISE_0 e ISE_1 in cui i pesi sono dati dalla lunghezza degli intervalli in cui il coefficiente funzionale è realmente nullo o diverso da zero, cioè

$$ISE(\hat{\beta}_j(t)) = ISE_0(\hat{\beta}_j(t))\mu(\mathcal{T}_{j,0}) + ISE_1(\hat{\beta}_j(t))\mu(\mathcal{T} - \mathcal{T}_{j,0}) \quad j = 1, 2.$$

Infine, se si vogliono considerare delle metriche generali relative alle stime di entrambi i coefficienti funzionali, verranno utilizzate semplicemente le somme delle metriche appena descritte calcolate separatamente per i due coefficienti.

4.3 Risultati

Vengono ora mostrati e commentati i risultati ottenuti nello studio di simulazione sopra descritto. Nella Tabella 4.2 sono riportate le medie Monte Carlo delle metriche ISE_0 e ISE_1 calcolate su $R = 50$ simulazioni effettuate per ogni combinazione considerata dei parametri dello studio. Per visualizzare l'intera distribuzione empirica delle stime Monte Carlo ottenute, si osservino invece i *boxplot* disegnati per ciascun caso e per ciascuno scenario, che sono riportati nelle Tabelle B.1 e B.2 in Appendice B. La tabella qui riportata mostra i risultati ottenuti per i due coefficienti di regressione funzionali considerati insieme. I risultati specifici riferiti a $\beta_1(t)$ e $\beta_2(t)$ separatamente sono invece riportati nella Tabella B.1 in Appendice e si può notare che sono entrambi in linea con l'andamento delle medie complessive. Nelle tabelle e nei grafici, il modello di regressione con sparsità condivisa è denominato con $LSFR_2$ mentre l'etichetta $LSFR_1$ è associata al modello localmente sparso di Bernardi et al. (2022).

TABELLA 4.2: Medie Monte Carlo ($\times 10^2$) relative alle metriche di valutazione riferite a entrambi i coefficienti funzionali insieme.

	<i>LSFR</i> ₂		<i>LSFR</i> ₁		<i>Lasso</i>		<i>Ridge</i>		<i>Elastic Net</i>	
	<i>ISE</i> ₀	<i>ISE</i> ₁	<i>ISE</i> ₀	<i>ISE</i> ₁	<i>ISE</i> ₀	<i>ISE</i> ₁	<i>ISE</i> ₀	<i>ISE</i> ₁	<i>ISE</i> ₀	<i>ISE</i> ₁
Caso 1										
S1	0.16	4.13	0.20	4.32	0.26	6.25	0.43	4.27	0.29	5.29
S2	0.24	7.01	0.36	7.62	0.44	9.57	0.53	6.94	0.41	8.52
S3	0.08	2.32	0.09	2.28	0.13	3.17	0.18	2.54	0.15	2.89
Caso 2										
S1	0.35	4.40	0.37	4.36	0.46	5.99	0.69	4.44	0.47	5.08
S2	0.68	7.71	0.83	8.01	0.90	9.86	1.04	7.48	0.84	8.70
S3	0.09	1.76	0.08	1.68	0.09	2.27	0.20	1.91	0.10	1.91
Caso 3										
S1	0.35	3.48	0.29	3.46	0.47	6.11	0.65	3.80	0.48	4.89
S2	0.62	6.03	0.61	6.31	0.94	9.09	1.03	6.54	0.99	7.77
S3	0.11	1.84	0.08	1.75	0.22	3.24	0.20	1.87	0.20	2.56

Osservando la Tabella 4.2, si può notare che i valori medi di ISE_0 ottenuti con i due modelli localmente sparsi $LSFR_1$ e $LSFR_2$, sono sistematicamente più bassi rispetto a quelli ottenuti con i modelli concorrenti. Negli scenari considerati, quindi, i modelli di regressione lineari con struttura di sparsità locale risultano essere empiricamente i migliori nell'identificare le zone in cui i due coefficienti funzionali sono realmente nulli. Anche per quanto riguarda l' ISE_1 si può notare che spesso esso è più basso nei due modelli che impongono una sparsità strutturata, ma vi sono alcuni casi in cui il modello migliore risulta essere quello con penalizzazione *Ridge*. Nelle zone in cui i coefficienti sono delle curve lisce, quindi, le stime possono essere meno accurate di quelle che si ottengono con un metodo di regolarizzazione che non fa selezione delle variabili. È noto, infatti, che i metodi che fanno selezione delle variabili producono stime maggiormente distorte per i coefficienti elevati. Come è stato dimostrato da Meinshausen & Bühlmann (2006) esiste un conflitto tra selezione delle variabili consistente e previsione ottimale conseguente a una stima corretta dei coefficienti di regressione. Il modello con penalizzazione *Ridge* non risente del problema e questo potrebbe essere il motivo per cui l'errore calcolato nelle zone in cui i coefficienti sono diversi da zero è talvolta più basso. I modelli *LSFR*, infatti, operano una selezione delle variabili ed è quindi ragionevole-

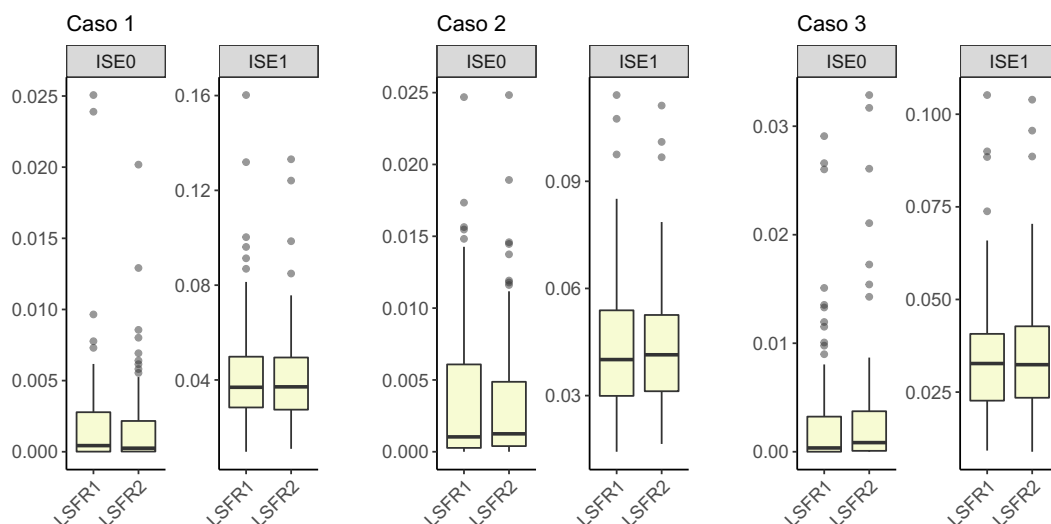


FIGURA 4.2: Distribuzione empirica di ISE_0 e ISE_1 ottenuta con il modello localmente sparso $LSFR_1$ e il modello con sparsità condivisa $LSFR_2$.

le aspettarsi che la stima dei coefficienti più grandi possa essere distorta. Ad ogni modo, questi modelli sembrano comunque risentire meno del problema rispetto al *Lasso*, che mostra valori di ISE_1 più grandi degli altri modelli in tutti gli scenari di simulazione.

Per quanto riguarda il confronto tra i modelli $LSFR_1$ ed $LSFR_2$, per una visione più completa si possono guardare anche i grafici in Figura 4.2, che mostrano le intere distribuzioni empiriche delle due metriche di valutazione ottenute con i due modelli per ciascuna delle tre forme considerate dei coefficienti funzionali. Per una visualizzazione più sintetica sono riportati solo i risultati ottenuti nello scenario di simulazione *standard* (S1), in cui la numerosità campionaria è pari a 50 e il rapporto segnale-rumore è 4. Per un confronto relativo anche agli altri scenari di simulazione si rimanda all'Appendice B.

Dai grafici in Figura 4.2 si evince che i risultati ottenuti con i due modelli localmente sparsi sono molto simili tra loro. Si può notare però che l'errore di stima del modello con sparsità condivisa calcolato nelle regioni realmente nulle

è leggermente più basso quando nei coefficienti di regressione funzionali vi è una evidente struttura di sparsità condivisa, come nel Caso 1. Quindi, in questo contesto, il modello $LSFR_2$ sembra essere quello che riesce meglio a cogliere la corretta struttura di sparsità dei coefficienti funzionali, mantenendo comunque una buona stima anche nelle regioni in cui essi sono delle curve lisce. Nel caso in cui, invece, il *pattern* di sparsità sia condiviso ma non in modo netto, non sembra esserci un vantaggio nell'utilizzo nel modello $LSFR_2$. Nelle zone agli estremi del dominio in cui uno dei due coefficienti funzionali è prossimo allo zero, le stime ottenute con il modello proposto nell'elaborato potrebbero infatti essere più distorte di quelle ottenute con il modello $LSFR_1$ a causa della diversa scala dei coefficienti delle basi che vengono penalizzati insieme. Infine, nel Caso 3, in cui la struttura di sparsità è diversa, come era ragionevole aspettarsi, con il modello proposto nel presente elaborato si ottengono risultati leggermente peggiori rispetto a quelli forniti dallo stesso modello che però non permette che le zone nulle dei due coefficienti siano le stesse.

Per un confronto più generale e diretto del metodo proposto con i quattro modelli concorrenti si può osservare la Figura 4.3. Nel grafico sono riportati, per ciascuna diversa forma assunta per i coefficienti, i *boxplot* dell'*ISE* relativo, calcolato come il rapporto tra l'errore quadratico integrato ottenuto con i modelli concorrenti e quello ottenuto con il modello $LSFR_2$. Anche in questo caso, si mostrano solo i risultati complessivi dei due coefficienti di regressione, relativi unicamente allo scenario di simulazione *standard*. Fatta eccezione per il modello $LSFR_1$, il terzo quantile dell'*ISE* relativo è quasi sempre a ridosso o al di sopra dell'uno, ad indicare che in più del 75% delle simulazioni il modello che permette una struttura di sparsità condivisa ha fornito dei risultati migliori rispetto ai modelli con penalizzazione *Lasso*, *Ridge* o *Elastic Net*. Si può notare in modo ancora più evidente di quanto si osservava nella Tabella 4.2, che i risultati ottenuti con il modello proposto sono nettamente più soddisfacenti di quelli del *Lasso* e di poco migliori rispetto alla *Ridge*, per i motivi sopra descritti.

Infine, si nota in maniera immediata che nel Caso 1 sembra che il modello introdotto nel presente elaborato sia leggermente migliore rispetto a quello che

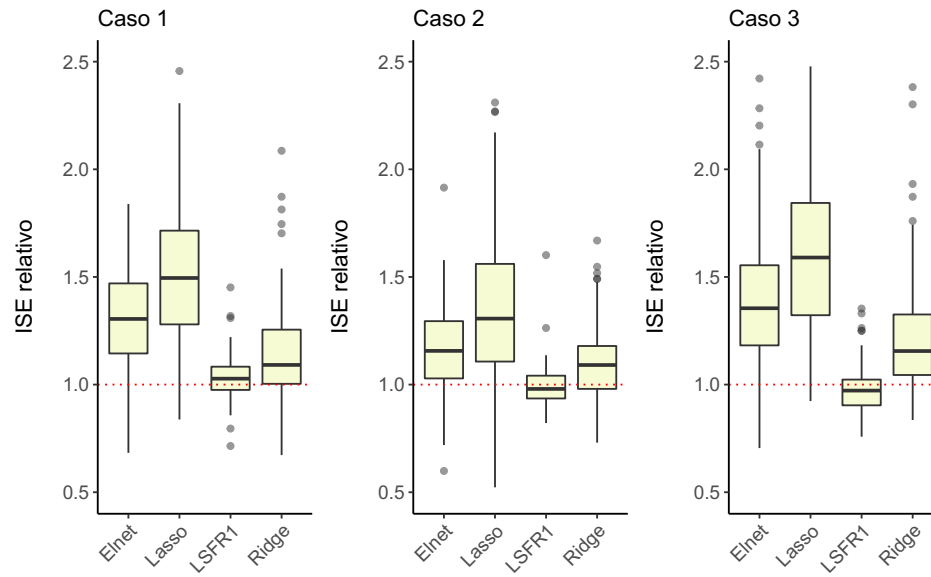


FIGURA 4.3: *ISE* relativo calcolato nello scenario *standard* ($n = 50$, $SNR = 4$) per i tre diversi casi.

non permette sparsità condivisa visto che la linea mediana è sopra all'uno, mentre il contrario accade per i Casi 2 e 3.

Si può quindi concludere che il modello di regressione funzionale con struttura di sparsità condivisa è risultato essere empiricamente il migliore rispetto ai modelli che non permettono selezione delle variabili o che ammettono solo sparsità non strutturata. È emerso inoltre, che nei casi in cui vi è effettivamente lo stesso *pattern* di sparsità nei due coefficienti funzionali, tenerne conto talvolta migliora, se pur di poco, l'accuratezza delle stime.

Capitolo 5

Analisi di alcuni dati meteorologici

In questo capitolo viene applicato il modello di regressione lineare funzionale con sparsità condivisa a un problema statistico reale. Verranno prima di tutto descritti il problema oggetto di studio e i dati che sono stati analizzati per affrontarlo. In seguito, dopo l'analisi esplorativa e il lisciamiento dei dati funzionali, verranno esposti i risultati ottenuti.

5.1 Descrizione del problema e dei dati

Il problema che ci si propone di affrontare è di carattere meteorologico e consiste in uno studio sulla relazione tra la temperatura e le precipitazioni. In particolare, per alcune centraline disposte in Veneto, si vuole prevedere la quantità totale di pioggia caduta nell'anno 2021, usando le serie storiche dell'andamento delle temperature minime e della loro velocità di cambiamento durante tutto il corso dell'anno. A tale fine, si costruisce un modello di regressione funzionale in cui la variabile risposta è il logaritmo della somma cumulata dei millimetri di pioggia caduti in tutti i giorni dell'anno e le due covariate funzionali sono le curve della temperatura minima giornaliera e le loro derivate.

I dati utilizzati sono stati reperiti nel sito dell'ARPAV (2022), l'Agenzia Regionale per la Prevenzione e la Protezione Ambientale del Veneto, e sono stati rilevati e registrati automaticamente da dei sensori posti in varie centraline presenti nelle

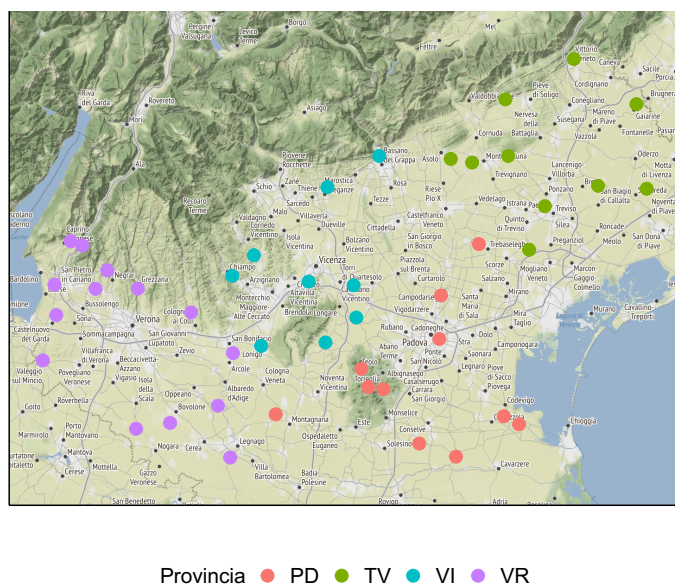


FIGURA 5.1: Collocazione geografica delle 44 centraline analizzate.

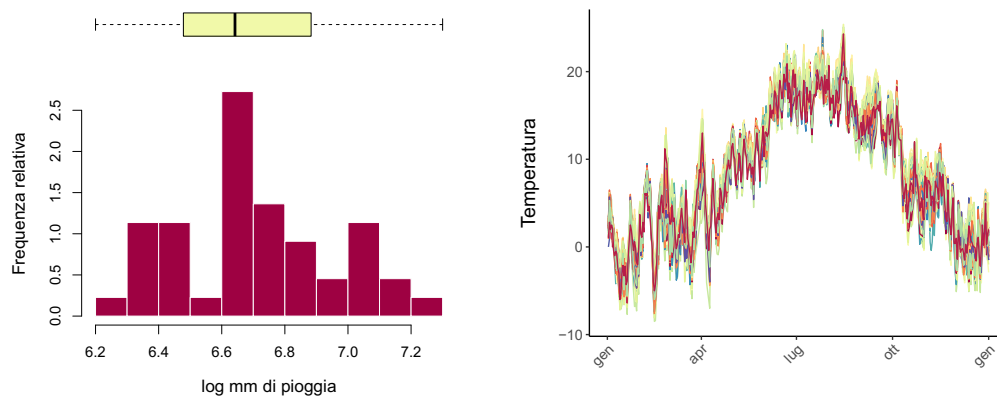
province di Padova, Treviso, Verona e Vicenza. I dati riportati nel sito sono stati anche sottoposti a un accurato processo di validazione da parte dell'ARPAV stesso. Per l'analisi sono state prese in considerazione solo le centraline poste a meno di 300 metri di quota al fine di concentrarsi sulle zone di pianura perché è noto che nelle zone montane e pedemontane la piovosità è influenzata anche da fattori differenti rispetto a quelli che sono determinanti nelle zone a bassa altitudine. Saldando sui rilievi, infatti, diminuiscono la temperatura e la pressione e la piovosità in genere aumenta fino a una certa quota per poi lasciare spazio alle precipitazioni nevose sui rilievi più alti. Per questo motivo si è scelto di focalizzare l'analisi sulle 44 centraline situate in pianura. Nella Figura 5.1 si può vedere nel dettaglio la collocazione geografica delle centraline considerate, che risultano essere disposte più o meno uniformemente su gran parte della pianura padano-Veneta.

5.2 Analisi esplorativa e preparazione dei dati

Come anticipato, i dati che vengono utilizzati nella costruzione del modello di regressione funzionale sono il numero totale di millimetri di pioggia caduti nel 2021

espressi in scala logaritmica e l'intera serie storica giornaliera della temperatura minima in gradi centigradi nelle 44 stazioni analizzate. Nella Figura 5.2 sono riportate delle analisi esplorative relative alle due variabili in questione. Nel pannello (a) vi sono dei grafici che mostrano la distribuzione empirica del logaritmo della quantità di pioggia caduta, che risulta essere unimodale. Nel pannello (b), invece, vi sono le 44 serie storiche della temperatura minima. Questa presenta un andamento medio circa crescente fino al mese di luglio e poi decrescente. Le serie, inoltre, mostrano numerose oscillazioni di breve periodo, che le rendono piuttosto irregolari.

I dati relativi alla temperatura verranno considerati come dati di tipo funzionale quindi è necessario per ciascuna unità statistica convertire in una funzione i punti osservati a cadenza giornaliera. Si esegue quindi serie per serie il lisciamiento con l'obiettivo anche di eliminare le fluttuazioni di breve periodo presenti nei dati e ottenere così funzioni lisce. Visto che si tratta di dati annuali periodici si utilizza un sistema di basi di Fourier di periodo pari a 365. Inoltre, dal momento che oltre



(a) Istogramma di frequenza del logaritmo dei mm totali di pioggia caduti. (b) Andamento della temperatura minima.

FIGURA 5.2: Analisi esplorativa della variabile scalare relativa alla quantità di pioggia caduta nel 2021 e delle serie storiche della temperatura minima nello stesso anno relative a ciascuna delle 44 centraline.

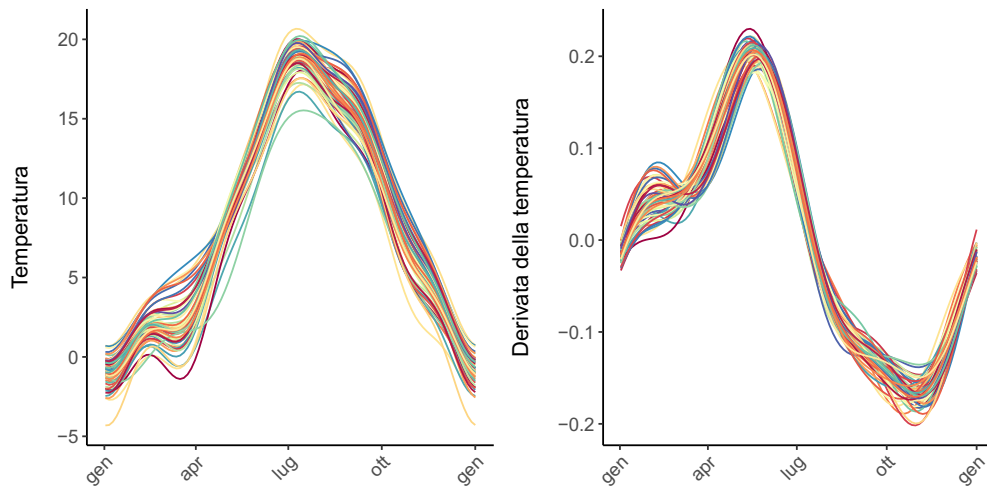


FIGURA 5.3: Curve lisce della temperatura minima e della sua derivata.

all'andamento della temperatura verrà utilizzata nelle analisi anche la sua derivata, occorre tenerne conto in fase di scelta dei parametri delle basi, in modo che anche la derivata risulti una curva liscia. Si utilizza allora un sistema di 7 funzioni di base e l'operatore differenziale usato nella penalizzazione della funzione obiettivo è l'operatore di accelerazione armonica, definito in Appendice A.3. Dopo aver stimato il modello di lisciamo con sistemi di funzioni di diverse dimensioni, da un'analisi visiva è emerso che l'insieme di 7 funzioni di Fourier fosse il giusto compromesso per ottenere curve lisce e non perdere troppa informazione. Nella Figura 5.3 sono riportate le curve della temperatura minima e della sua derivata risultate dall'operazione di lisciamo appena descritta. Le curve sono state poi normalizzate, in modo che il *range* di variabilità di tutte le funzioni sia uniformato all'intervallo $[0, 1]$.

I due insiemi di funzioni lisce ottenuti verranno ora utilizzati come covariate di un modello di regressione lineare funzionale. Il modello di regressione funzionale con sparsità condivisa potrebbe risultare particolarmente adeguato perché è ragionevole aspettarsi che la temperatura e la velocità con cui essa cambia abbiano un impatto sulla quantità totale di pioggia caduta solo negli intervalli temporali riferiti ai periodi che effettivamente sono stati piovosi. Il grafico in Figura 5.4 mo-

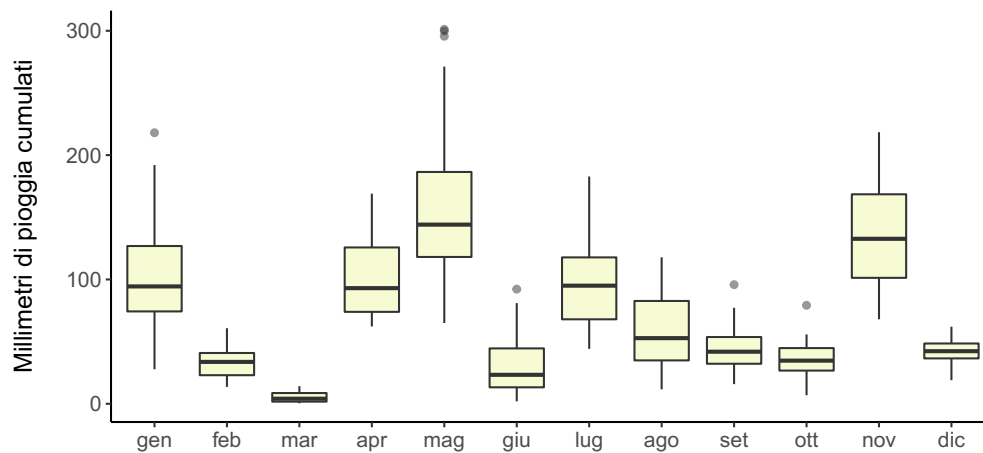


FIGURA 5.4: Distribuzione della somma cumulata dei millimetri di pioggia caduti nelle 44 centraline, condizionata ai dodici mesi dell'anno.

stra le distribuzioni della somma dei millimetri di pioggia rilevati nelle centraline condizionate al mese dell'anno. Si può notare che il livello mediano delle distribuzioni è piuttosto variabile e vi sono dei mesi in cui la quantità di pioggia rilevata è talmente bassa che probabilmente non ha piovuto quasi mai. È quindi ragionevole pensare che vi siano degli intervalli del dominio in cui la variabile risposta non è influenzata dalla temperatura e dalla sua derivata.

5.3 Stima del modello di regressione funzionale con sparsità condivisa

Sui dati normalizzati è stato quindi stimato il modello di regressione funzionale con sparsità condivisa che è stato presentato nel Capitolo 3.

Il coefficiente di regressione funzionale è stato espresso tramite un insieme di *B-splines* di ordine 4 e di dimensione pari a 30. Il numero di funzioni di base da utilizzare è stato scelto in modo tale da far sì che il numero di gruppi sulla base dei quali è costruita la funzione di penalizzazione fosse abbastanza elevato in modo da permettere un maggior numero di *pattern* di sparsità differenti. D'altra parte però,

un sistema di basi di dimensione elevata comporta una maggiore complessità del modello e un maggior onere computazionale perché aumenta il numero di parametri da stimare. È quindi stato scelto un valore che costituisse un compromesso tra flessibilità e complessità del modello e a seguito di diverse prove con insiemi di numerosità pari a 20, 25, 30 e 35, il sistema con 30 *B-splines* è risultato essere il più adeguato.

Il parametro di regolazione λ è stato invece scelto con il metodo della convalida incrociata a cinque *fold* e ad ogni iterazione è stato stimato il modello con sparsità condivisa per una griglia di 20 diversi valori del parametro. Il grafico in Figura 5.5 mostra l'andamento dell'errore quadratico medio di convalida incrociata calcolato per ciascuno valore di λ considerato. I limiti superiore e inferiore delle relative bande di confidenza valide punto a punto sono stati calcolati rispettivamente sommando e sottraendo la deviazione standard.

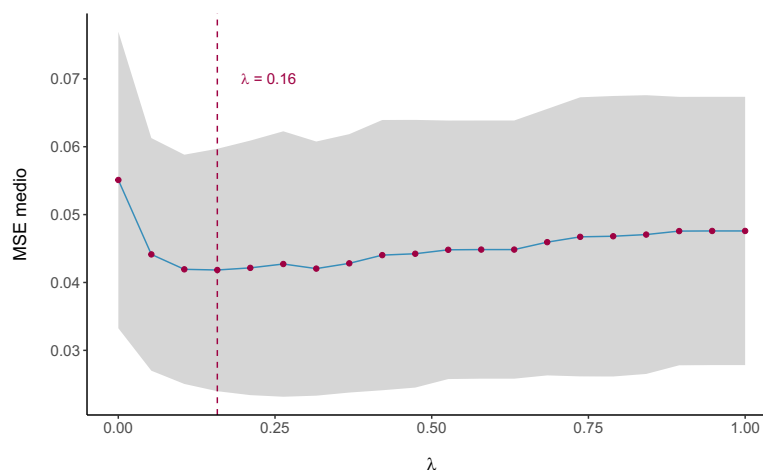


FIGURA 5.5: MSE medio di convalida incrociata con relative bande di confidenza per 20 diversi valori del parametro di regolazione λ .

La scelta è quindi ricaduta sul valore del parametro di regolazione che minimizza l'errore medio di convalida incrociata, e come si può vedere dalla figura, questo è risultato essere pari a 0.16.

Per tener conto del livello medio della variabile risposta, nel modello è stata inserita anche l'intercetta. Con i valori scelti per i parametri ($M = 30, \lambda = 0.16$),

è stato quindi ri-stimato il modello usando l'intero *dataset*.

Oltre al modello con sparsità condivisa, sono stati stimati anche altri due modelli di regressione funzionale in modo da valutare più accuratamente l'adeguatezza del modello proposto confrontandone i risultati. A tale fine è stato quindi stimato il modello di regressione funzionale localmente sparso con una sola covariata e lo stesso modello con le due covariate funzionali e la penalizzazione che non tiene conto dei gruppi che inducono sparsità condivisa. Per entrambi i modelli, in modo analogo a quanto effettuato in precedenza, i coefficienti funzionali sono stati rappresentati con 30 *B-splines* di ordine 4 e il parametro di lisciamiento è stato scelto con convalida incrociata, in cui la suddivisione in 5 gruppi è la stessa che era stata utilizzata per il modello precedente. I modelli di regressione funzionale sono stati poi confrontati usando la media dell'errore quadratico medio di previsione ottenuto nei cinque *fold* della convalida incrociata.

I risultati del modello con sparsità condivisa esposto in questo elaborato sono riportati e commentati nel paragrafo seguente. A supporto della tesi secondo cui il modello risulta essere adeguato per il problema in esame, sono riportati anche alcuni dei risultati ottenuti coi due modelli concorrenti considerati.

5.4 Risultati

I due coefficienti funzionali stimati con i dati normalizzati dal modello di regressione con sparsità condivisa sono rappresentati nel grafico in Figura 5.6.

Il primo coefficiente di regressione funzionale, associato al livello della temperatura, è negativo o nullo in tutto il suo dominio. Nello specifico, esso è negativo in tutta la prima parte dell'anno, fino al 29 luglio, ad eccezione di un piccolo intervallo temporale a fine marzo in cui il coefficiente è prossimo allo zero. Sembra quindi che nel 2021, una diminuzione della temperatura nei primi sette mesi dell'anno avrebbe comportato un aumento della quantità di pioggia totale caduta.

Il coefficiente funzionale associato alla derivata della temperatura è invece positivo nei mesi di gennaio, febbraio, giugno, luglio e nei primi 10 giorni di agosto, a seguito dei quali la stima è pari a zero. Nei mesi di aprile e maggio, invece, il coef-

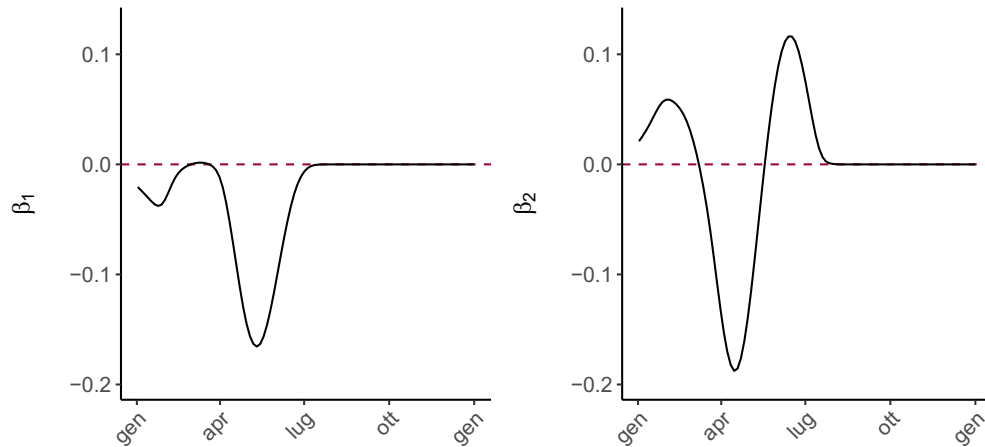


FIGURA 5.6: Coefficienti funzionali stimati con il modello di regressione funzionale con sparsità condivisa espressi nella scala normalizzata.

ficiente è negativo. Sulla base della stima prodotta dal modello emerge quindi che un aumento della velocità di cambiamento nella temperatura nei mesi di gennaio, febbraio, giugno e luglio del 2021 avrebbe portato a un incremento della quantità di pioggia caduta. Al contrario, se la temperatura fosse cambiata più velocemente ad aprile e maggio si sarebbe osservata una diminuzione della pioggia totale.

Nell'anno 2021, sembra quindi che da agosto in poi sia la temperatura sia le velocità con cui essa cambia non abbiano influenzato in modo apprezzabile la quantità totale di pioggia rilevata nelle 44 centraline della pianura padano-Veneta. I coefficienti funzionali stimati presentano quindi effettivamente una struttura di sparsità condivisa, il che ha permesso di interpretare in modo più semplice e chiaro i risultati del modello di regressione funzionale.

Per valutare che la struttura di sparsità condivisa sia realmente sottesa al fenomeno in esame il modello è stato confrontato in termini previsivi con il modello di regressione funzionale localmente sparso che non permette sparsità condivisa. Nella Tabella 5.1 sono riportate le misure medie di errore e il valore ottimo del parametro di regolazione associati a ciascuno dei modelli stimati. È emerso che l'errore medio del modello che permette la medesima struttura di sparsità nei coef-

ficienti funzionali è leggermente più basso di quello del secondo modello stimato. Consentire in fase di stima che la struttura di sparsità sia la stessa nei due coefficienti funzionali, oltre ad agevolare l'interpretazione del fenomeno, ha quindi anche permesso di ridurre un po' l'errore medio di previsione.

Per poter poi interpretare i risultati del modello anche a livello quantitativo occorre retro-trasformare i coefficienti stimati usando la Formula 3.5, che permette di riportare alla scala originale le stime dei coefficienti ottenute usando i dati normalizzati. Le stime dei coefficienti funzionali nella scala originale sono visibili nella Figura 5.7. Dai due diversi ordini di grandezza dei coefficienti stimati si deduce che la velocità con cui cambia la temperatura sembra avere un impatto sulla pioggia decisamente più importante rispetto al livello della temperatura stessa. Nonostante la notevole differenza di scala, anche l'effetto del livello della temperatura è comunque utile a prevedere la somma dei millimetri di pioggia caduti espressa in scala logaritmica. Come si può notare dalla Tabella 5.1, infatti, il modello localmente sparso che utilizza come unica covariata la derivata della temperatura ha portato a un errore medio di convalida incrociata superiore rispetto a quello dei modelli che tengono conto anche del livello della temperatura.

TABELLA 5.1: Valori medi di *Mean Squared Error* e valori ottimi del parametro di regolazione per i tre modelli stimati.

Modello	λ ottimo	MSE medio
<i>LSFR</i> ₂ con temperatura e sua derivata	0.16	0.0418
<i>LSFR</i> ₁ con temperatura e sua derivata	0.16	0.0430
<i>LSFR</i> con derivata della temperatura	0.14	0.0465

I confronti effettuati tra i diversi modelli di regressione funzionale stimati hanno il limite di non essere stati condotti sulla base di rigorosi test statistici con associata una misura di significatività. Pur tenendo conto di questo limite, i confronti avevano l'obiettivo di confermare o smentire l'intuizione iniziale secondo cui, nel fenomeno oggetto di studio, entrambe le covariate funzionali considerate fossero effettivamente in relazione con la variabile risposta e che tale relazione esistesse

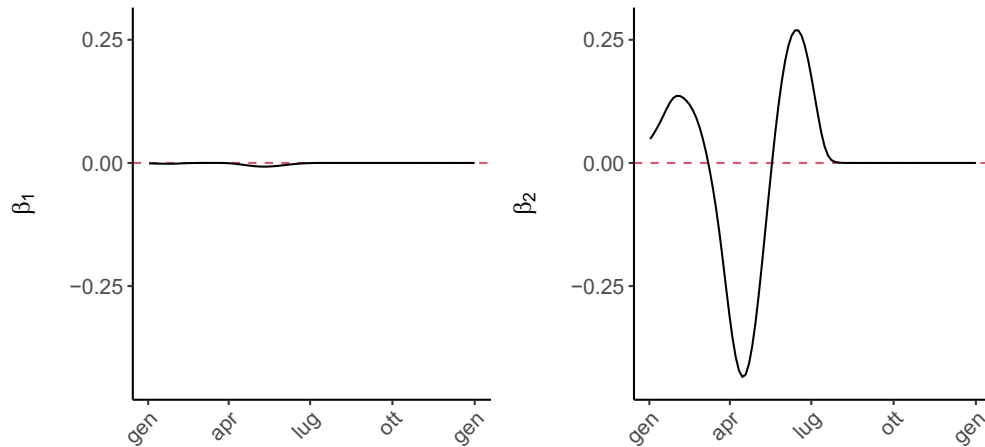


FIGURA 5.7: Coefficienti funzionali stimati con il modello di regressione funzionale con sparsità condivisa espressi nella scala originale.

solo in determinate regioni comuni del loro dominio. Alla luce dei risultati ottenuti, si può quindi concludere che il modello di regressione funzionale localmente sparso con struttura di sparsità condivisa sembra essere adeguato a cogliere le peculiarità del fenomeno oggetto di studio, favorendo così un'agevole interpretazione e garantendo comunque anche buone capacità previsive.

Conclusioni

Lo scopo di questo elaborato è stato quello di proporre un particolare metodo di stima per il modello di regressione lineare funzionale da utilizzare nel caso in cui, nel problema oggetto di studio, due variabili esplicative di tipo funzionale vengano impiegate per la previsione della variabile risposta scalare. La funzione di penalizzazione proposta è stata ideata in modo da permettere che la struttura di sparsità possa essere condivisa tra i due coefficienti di regressione stimati, presupponendo che nel fenomeno in esame le due covariate funzionali siano in relazione con la variabile risposta solo in determinate regioni comuni del loro dominio. La penalizzazione costruita è stata quindi utilizzata in fase di stima nel criterio dei minimi quadrati penalizzati. L'obiettivo è quindi stato quello di proporre uno strumento che, se applicato a un problema statistico con queste caratteristiche, permetta di ottenere delle stime facilmente interpretabili pur mantenendo anche buone capacità previsive.

Si è visto che in un problema statistico che presenta le peculiarità per le quali è stata costruita la penalizzazione proposta, le stime ottenute usando tale penalizzazione sono risultate essere più accurate rispetto a quelle ricavate tramite il modello stimato con le penalizzazioni *Ridge*, *Lasso* o *Elastic net*. Inoltre, è stato empiricamente dimostrato che il modello proposto è in grado di identificare le zone di sparsità dei coefficienti funzionali in modo più preciso rispetto al modello di regressione funzionale localmente sparso di Bernardi et al. (2022), nel caso in cui la struttura di sparsità sia realmente condivisa tra i due coefficienti funzionali.

Un aspetto che sarebbe utile approfondire per migliorare i risultati del modello di regressione con sparsità condivisa riguarda l'ottimizzazione numerica della

funzione obiettivo. Nel presente elaborato è stato utilizzato l'algoritmo di ottimizzazione numerica implementato nella funzione `nlinb`. Per ridurre i tempi di calcolo e controllare in modo più accorto la convergenza, si potrebbe utilizzare un algoritmo di ottimizzazione iterativa di tipo MM (*Majorize-Minimization*), che sfrutta una funzione surrogata più semplice per minimizzare la funzione obiettivo.

Appendice A

Strumenti di algebra e analisi matematica

A.1 Operatore differenziale

Gli operatori differenziali sono delle trasformazioni dell'operatore di derivazione e quelli più largamente diffusi sono gli operatori differenziali lineari, i quali, agendo come trasformazioni lineari, conservano le operazioni di somma e prodotto. Si definisce l'operatore differenziale lineare L di ordine n come

$$L = w_n(x)D_x^n + w_{n-1}D_x^{(n-1)} + \dots + w_1(x)D_x + w_0(x)$$

dove $n \in \mathbb{N}$, $w_j(x)$ con $j = 0, \dots, n$ sono funzioni peso differenziabili e

$$D_x^{(k)} = \frac{d^k}{dx^k}$$

è l'operatore di derivazione di ordine k .

L'operatore differenziale più semplice è la derivata. Alcuni degli operatori differenziali maggiormente utilizzati sono l'accelerazione e l'operatore di accelerazione armonica, che vengono spesso considerati come indici di irregolarità delle curve rispettivamente non periodiche e periodiche. L'accelerazione è pari alla derivata seconda, la quale costituisce una misura di convessità/concavità della curva a cui

è applicata. L'operatore di accelerazione armonica, invece, è definito come

$$L = w^2 D_x + D_x^3,$$

dove w è in relazione con il periodo T della curva $f(x)$ di cui si intende misurare il grado di irregolarità ($w = 2\pi/T$). Si può dimostrare che l'operatore di accelerazione armonica, applicato alla curva periodica $f(x)$, costituisce effettivamente una misura di irregolarità di tale curva. Infatti, tale operatore è nullo se applicato a una funzione regolare di periodo T definita come

$$x(t) = c_0 + a_1 \sin(wt) + b_1 \cos(wt),$$

dove c_0 , a_1 e b_1 sono i parametri reali della curva.

Per una trattazione più approfondita si vedano ad esempio Lanczos (1996) o Euler (2012).

A.2 Formula di Cox-De Boor

La formula ricorsiva di Cox-De Boor viene utilizzata per la costruzione di funzioni di base di tipo *B-splines*.

Siano $t_0, t_1, t_2, \dots, t_K$ i nodi interni e $t_0 < t_1$ e $t_{K+1} > t_K$ i due nodi esterni. Si definisce un ulteriore insieme di nodi composto da

- $\tau_1 \leq \tau_2 \leq \dots, \leq \tau_M \leq t_0$
- $\tau_{M+k} = t_k$ con $j = k, \dots, K$
- $t_{K+1} \leq \tau_{M+K+1} \leq \dots \leq \tau_{K+2M}$

dove i primi e gli ultimi M nodi possono essere uguali tra loro e uguali ai due nodi esterni.

Data la sequenza di nodi $\{\tau_j\}^{j=1, \dots, K+2M}$, le *B-splines* di ordine uno sono definite da

$$\phi_j^1(t) = \begin{cases} 1 & \text{se } \tau_j \leq t \leq \tau_{j+1} \\ 0 & \text{altrimenti} \end{cases}$$

per $j = 1, \dots, K + 2M - 1$. Le *B-splines* di generico ordine superiore pari a m sono definite ricorsivamente come

$$\phi_j^{(m)}(t) = \frac{t - \tau_j}{\tau_{j+K-1} - \tau_j} \phi_j^{(m-1)}(t) - \frac{t - \tau_{j+K}}{\tau_{j+K} - \tau_{j+1}} \phi_{j+1}^{(m-1)}(t)$$

con $j = 1, \dots, K + 2M - m$.

A.3 Serie di Fourier

Lo sviluppo in serie di Fourier è una rappresentazione di una funzione periodica $f(x)$ mediante una combinazione lineare delle funzioni sinusoidali

$$\begin{aligned} \phi_1(t) &= 1 \\ \phi_2(t) &= \sin(\omega t) \\ \phi_3(t) &= \cos(\omega t) \\ \phi_4(t) &= \sin(2\omega t) \\ \phi_5(t) &= \cos(2\omega t) \\ &\vdots \end{aligned} \tag{A.1}$$

dove la costante ω è in relazione con il periodo T della funzione $f(x)$ tramite la relazione $\omega = 2\pi/T$ (Cai & Hall, 2006). Come si può notare, dopo la prima funzione di base costante, le funzioni di base di Fourier sono organizzate in m coppie successive di funzioni seno e coseno. Le serie contengono solitamente entrambi gli elementi di ciascuna coppia, quindi il numero di funzioni di base è pari a $K = 2m + 1$.

La rappresentazione mediante lo sviluppo in serie di Fourier gode della proprietà di chiusura rispetto alla differenziazione: la derivata di un'espansione in serie di Fourier è ancora un'espansione in serie di Fourier.

A.4 Prodotto di Hadamard

Il prodotto di Hadamard, noto anche come prodotto *element-wise*, è un'operazione binaria che coinvolge due matrici di uguali dimensioni e ne restituisce una terza, di

dimensioni pari a quelle degli operandi. L'elemento (i, j) della matrice risultante è pari al prodotto degli elementi (i, j) delle due matrici di partenza.

Formalmente, per due matrici delle stesse dimensioni $A, B \in \mathbb{C}^{m \times n}$, il prodotto di Hadamard è una matrice $A \odot B \in \mathbb{C}^{m \times n}$, i cui coefficienti sono

$$(A \odot B)_{i,j} = (A)_{i,j} \times (B)_{i,j} \quad 1 \leq i \leq n, 1 \leq j \leq m.$$

Per matrici di dimensioni diverse, il prodotto di Hadamard non è definito.

Appendice B

Risultati aggiuntivi dello studio di simulazione

Sono di seguito riportati una tabella e dei grafici utili per visualizzare in modo più completo i risultati dello studio di simulazione.

TABELLA B.1: Medie Monte Carlo ($\times 10^2$) relative alle metriche di valutazione calcolate per i coefficienti funzionali $\beta_1(t)$ e $\beta_2(t)$.

		<i>LSFR₂</i>		<i>LSFR₁</i>		<i>Lasso</i>		<i>Ridge</i>		<i>Elastic Net</i>	
		<i>ISE₀</i>	<i>ISE₁</i>	<i>ISE₀</i>	<i>ISE₁</i>	<i>ISE₀</i>	<i>ISE₁</i>	<i>ISE₀</i>	<i>ISE₁</i>	<i>ISE₀</i>	<i>ISE₁</i>
Caso 1											
S1	$\beta_1(t)$	0.18	3.00	0.23	3.04	0.28	4.58	0.39	3.13	0.33	3.85
	$\beta_2(t)$	0.14	5.27	0.17	5.59	0.25	7.92	0.47	5.42	0.25	6.73
S2	$\beta_1(t)$	0.31	4.88	0.47	5.09	0.49	7.10	0.52	5.05	0.50	6.09
	$\beta_2(t)$	0.17	9.14	0.26	10.14	0.38	12.04	0.53	8.83	0.33	10.95
S3	$\beta_1(t)$	0.08	1.85	0.09	1.78	0.13	2.25	0.17	1.93	0.16	2.24
	$\beta_2(t)$	0.09	2.80	0.09	2.78	0.13	4.09	0.20	3.15	0.14	3.53
Caso 2											
S1	$\beta_1(t)$	0.37	3.97	0.39	3.78	0.48	5.41	0.64	3.93	0.51	4.77
	$\beta_2(t)$	0.33	4.83	0.35	4.95	0.44	6.57	0.74	4.95	0.44	5.38
S2	$\beta_1(t)$	0.76	6.35	0.98	6.49	1.01	8.33	1.07	6.17	0.96	7.31
	$\beta_2(t)$	0.60	9.07	0.67	9.53	0.79	11.40	1.00	8.79	0.72	10.09
S3	$\beta_1(t)$	0.08	1.83	0.07	1.66	0.08	2.22	0.19	1.79	0.09	2.07
	$\beta_2(t)$	0.10	1.70	0.10	1.70	0.10	2.32	0.21	2.03	0.11	1.74
Caso 3											
S1	$\beta_1(t)$	0.27	3.39	0.28	3.26	0.38	5.15	0.58	3.53	0.43	4.23
	$\beta_2(t)$	0.42	3.57	0.29	3.66	0.57	7.07	0.71	4.06	0.53	5.56
S2	$\beta_1(t)$	0.55	5.45	0.69	5.63	0.90	7.60	0.95	5.70	0.93	6.55
	$\beta_2(t)$	0.70	6.61	0.52	6.98	0.99	10.57	1.11	7.38	1.04	9.00
S3	$\beta_1(t)$	0.07	1.89	0.06	1.75	0.13	2.39	0.18	1.92	0.13	2.23
	$\beta_2(t)$	0.14	1.78	0.10	1.75	0.31	4.08	0.23	1.81	0.27	2.89

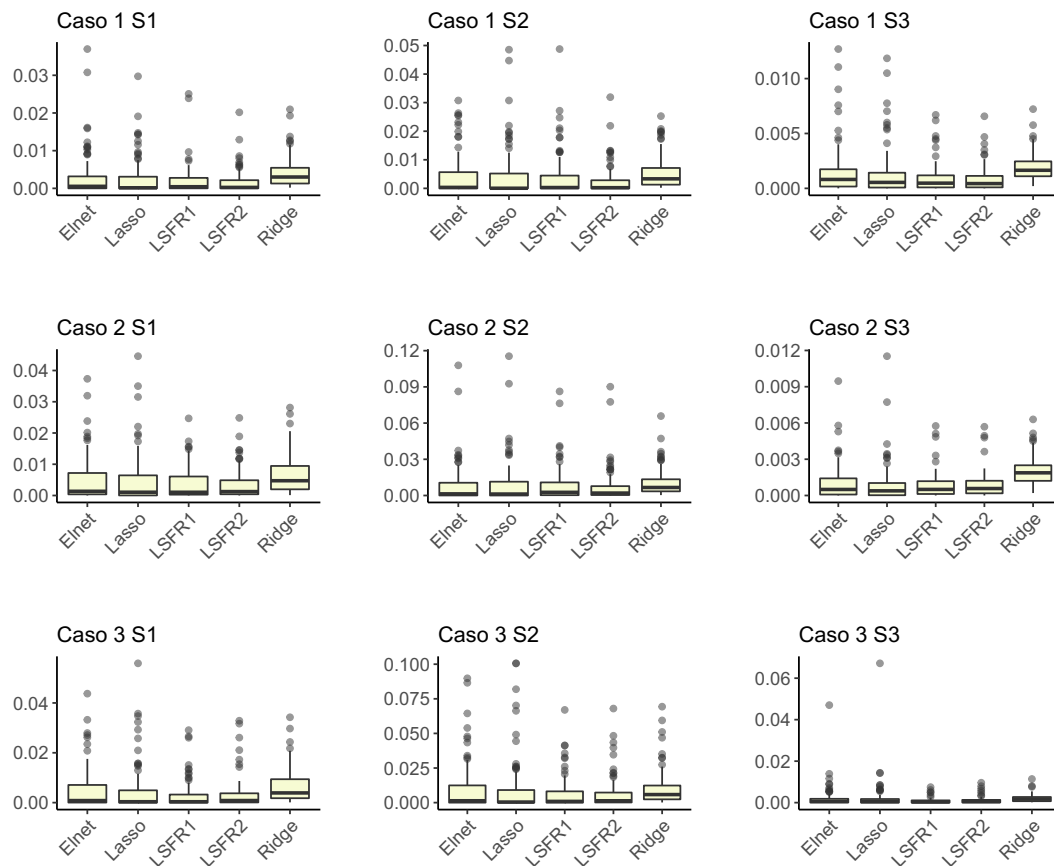


FIGURA B.1: Distribuzioni empiriche dei valori di ISE_0 riferiti a entrambi i coefficienti funzionali insieme, ottenuti nelle 50 simulazioni per ogni combinazione dei parametri dello studio.

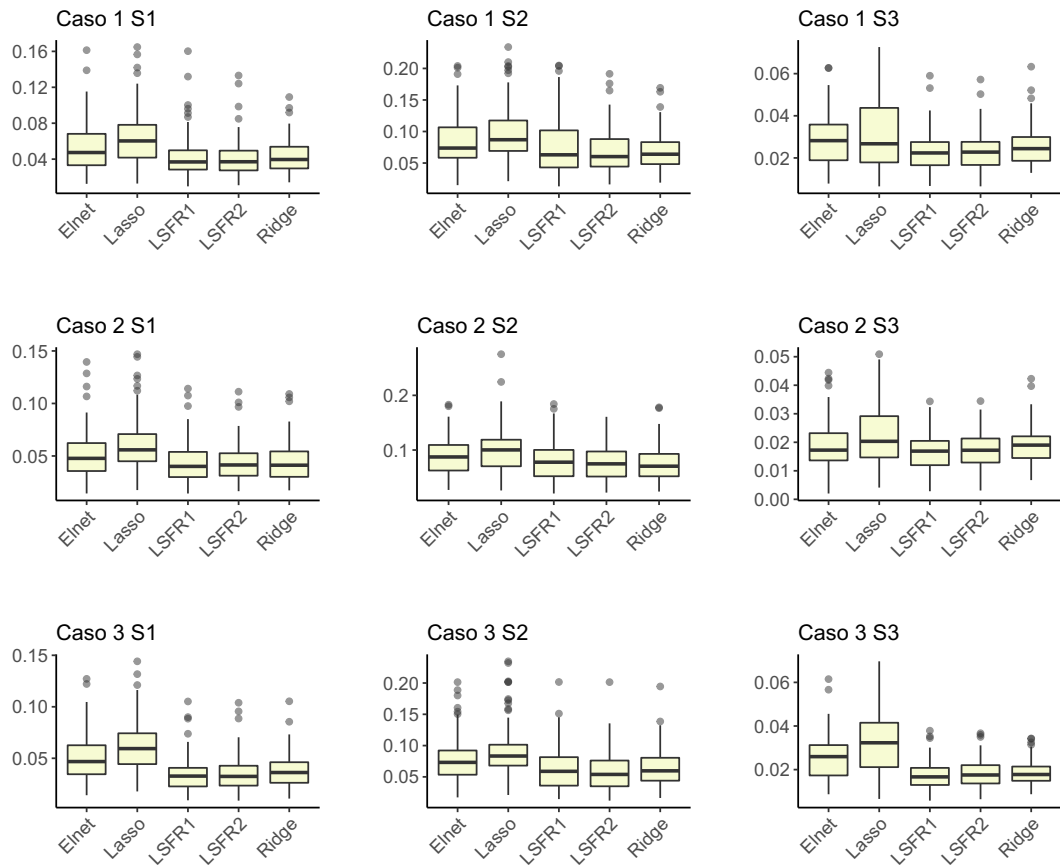


FIGURA B.2: Distribuzioni empiriche dei valori di ISE_1 riferiti a entrambi i coefficienti funzionali insieme, ottenuti nelle 50 simulazioni per ogni combinazione dei parametri dello studio.

Appendice C

Codice R utilizzato

Di seguito viene proposta una possibile implementazione di alcune funzioni utili per il calcolo della penalizzazione *Overlap Group lasso* e per la stima del modello di regressione funzionale localmente sparso con struttura di sparsità condivisa per uno specifico valore del parametro di regolazione.

CODICE C.1: Definizione dei gruppi.

```
getGroupMat_OVGLASS02x <- function(p, spl0rd = c(3, 4), share
  = TRUE) {
  "
  INPUT:
  p      Dimensione del sistema di basi con cui sono
         espressi i coefficienti funzionali
  spl0rd Ordine delle bsplines
  share  Flag per includere o meno i gruppi che premettono
         sparsita condivisa

  OUTPUT:
  Matrici di selezione vettorizzate
  "
  B <- (p - spl0rd + 1)
  grMat <- matrix(0, B, p)
  if ((spl0rd == 3) || (spl0rd == 4)) {
```

```

    for(i in 1:(p-splOrd+1)){
      grMat[i,(i:(i+splOrd-1))] <- 1
    }
  } else {
    cat("Error in getGroupMat_OVGLASSO, the spline order in
        not equat to 3 or 4!\n")
  }

  grMat2x <- matrix(0, nrow = 3*B, ncol = 2*p*q)
  grMat2x[1:B, 1:(p*q)] <- grMat
  grMat2x[(B+1):(2*B), (p*q+1):(2*p*q)] <- grMat
  if(share == TRUE){
    grMat2x[(2*B+1):(3*B), ] <- cbind(grMat, grMat)
  }
  grMat2x <- rbind(grMat2x, c(rep(1, p*q), rep(0, p*q)))
  grMat2x <- rbind(grMat2x, c(rep(0, p*q), rep(1, p*q)))
  return(grMat2x)
}

```

CODICE C.2: Calcolo della penalizzazione.

```

fs2s_penfun <- function(p, splOrd, regpars, share = TRUE) {
  "
  INPUT:
  p          Dimensione del sistema di basi con cui sono
             espressi i coefficienti funzionali
  splOrd     Ordine delle bsplines
  regpars    Vettore di lunghezza 2p con le stime dei
             coefficienti delle basi delle due covariate
  share      Flag per includere o meno i gruppi che premettono
             sparsita condivisa

  OUTPUT:
  Ammontare della penalizzazione Overlap Group lasso

```



```

"
B <- p - splOrd + 1
grMat <- getGroupMat_OVGLASS02x(p = p, splOrd = splOrd,
  share = share)
grMatWeights <- 1.0 / apply(grMat, 2, sum)
Cb <- t(apply(grMat, 1, function(x) x*grMatWeights))
norme_ <- apply(Cb, 1, function(x) sqrt(sum((x * regpars)
  ^2)))
norme <- fSet2GivenTolerance(x = norme_, n = B, toler = .
  Machine$double.eps)
return(sum(norme))
}

```

CODICE C.3: Stima del modello per un valore di λ .

```

fs2sreg1lambda <- function(Y, Z, M, splOrd, lambda, psi_INIT
  = NULL, beta.threshold.toler = 1e-4, maxiter = 100, share
  = TRUE, intercept = TRUE){
"
INPUT:
Y           Vettore coi valori della variabile
            risposta
Z           Matrice del disegno
M           Dimensione dell'insieme di basi con
            cui sono espressi i coefficienti
            funzionali
splOrd      Ordine delle bsplines
lambda      Valore del parametro di regolazione
psi_INIT    Inizializzazione per il vettore di
            coefficienti delle basi delle due
            covariate
beta.threshold.toler  Soglia minima di tolleranza sotto
            alla quale le stime vengono poste a
            zero

```

```
maxiter           Massimo numero di iterazioni per
                  nlminb
share             Flag per includere o meno i gruppi
                  che premettono sparsita condivisa
intercept         Flag per includere o meno
                  l'intercetta nel modello

OUTPUT:
stima del vettore di coefficienti delle basi delle due
  covariate
"

start.time <- Sys.time()
N <- NROW(Z)
# Model without intercept
-----
if(intercept == FALSE){
  if(is.null(psi_INIT)){
    psi_INIT <- rep(.Machine$double.eps, 2*M) + rnorm(2*M)
  }
  opt.fun <- function(psi){
    mse <- 0.5*sum((Y - Z %*% psi)^2)
    pen <- fs2s_penfun(p = M, spl0rd = spl0rd, regpars =
      psi, share = share)
    l <- mse + lambda * pen
    l
  }
  stima <- nlminb(psi_INIT, opt.fun, control = list(iter.
    max = maxiter))
  psi.hat_ <- stima$par
  psi.hat <- fthreshold(psi.hat_, length(psi.hat_), oler =
    beta.threshold.toler)
  intercept <- 0
```

```
end.time <- Sys.time()
}
# Model with intercept
-----
if(intercept == TRUE){
  if(is.null(psi_INIT)){
    psi_INIT <- rep(.Machine$double.eps, 2*M+1) + rnorm(2*M
      +1)
  }
  opt.fun <- function(psi){
    mse <- 0.5*sum((Y - psi[1] - Z %*% psi[-1])^2)
    pen <- fs2s_penfun(p = M, spl0rd = spl0rd, regpars =
      psi[-1], share = share)
    l <- mse + lambda * pen
    l
  }
  stima <- nlminb(psi_INIT, opt.fun, control = list(iter.
    max = maxiter))
  psi.hat_ <- stima$par[-1]
  psi.hat <- fthreshold(psi.hat_, length(psi.hat_), toler
    = beta.threshold.toler)
  intercept <- stima$par[1]
  end.time <- Sys.time()
}
# output
out <- list()
out$convergence <- ifelse(stima$convergence == 0, "
  successful convergence", "not successful convergence")
out$iter <- stima$iterations
out$loss <- stima$objective
out$elapsedTime <- difftime(end.time, start.time, units = "
  secs")[[1]]
out$psi.hat <- psi.hat
```

```
out$intercept <- intercept
return(out)
}
```

Bibliografia

- ARPAV (2022). Dati storici – arpa veneto. <https://www.arpa.veneto.it/arpavinforma/bollettini/dati-storici>. [Online; accessed 07-June-2022].
- AZZALINI, A. & SCARPA, B. (2009). *Analisi dei dati e data mining*. Springer.
- BATES, D. & VENABLES, B. (2011). *Regression Spline Functions and Classes*. R package version 4.2.1.
- BERNARDI, M., CANALE, A. & STEFANUCCI, M. (2022). Locally sparse function-on-function regression. *Journal of Computational and Graphical Statistics* doi:10.1080/10618600.2022.2130926.
- CAI, T. T. & HALL, P. (2006). Prediction in functional linear regression. *The Annals of Statistics* **34**, 2159–2179.
- CARDOT, H., FERRATY, F. & SARDA, P. (1999). Functional linear model. *Statistics & Probability Letters* **45**, 11–22.
- DE BOOR, C. (1978). *A practical guide to splines*, vol. 27. Springer-Verlag New York.
- EULER, N. (2012). Linear operators and the general solution of elementary linear ordinary differential equations. *CODEE Journal* **9**, 11.
- FRIEDMAN, J., HASTIE, T., TIBSHIRANI, R., BALASUBRAMANIAN, N., KENNETH, T., NOAH, S. & YANG, J. (2010). *Lasso and Elastic-Net Regularized Generalized Linear Models*. R package version 4.1.4.

- HALL, P., MÜLLER, H.-G. & WANG, J.-L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *The Annals of Statistics* **34**, 1493–1517.
- HAN, J., PEI, J. & TONG, H. (2022). *Data mining: concepts and techniques*. Morgan Kaufmann.
- HASTIE, T., TIBSHIRANI, R., FRIEDMAN, J. H. & FRIEDMAN, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction*, vol. 2. Springer.
- HASTIE, T., TIBSHIRANI, R. & WAINWRIGHT, M. (2015). *Statistical learning with sparsity*, vol. 143. Monographs on statistics and applied probability.
- HORVÁTH, L. & KOKOSZKA, P. (2012). *Inference for functional data with applications*, vol. 200. Springer Science & Business Media.
- JACOB, L., OBOZINSKI, G. & VERT, J.-P. (2009). Group lasso with overlap and graph lasso. In *Proceedings of the 26th annual international conference on machine learning*.
- JAMES, G. M., WANG, J. & ZHU, J. (2009). Functional linear regression that's interpretable. *The Annals of Statistics* **37**, 2083 – 2108. doi:10.1214/08-AOS641.
- JENATTON, R., AUDIBERT, J.-Y. & BACH, F. (2011). Structured variable selection with sparsity-inducing norms. *The Journal of Machine Learning Research* **12**, 2777–2824.
- LANCZOS, C. (1996). *Linear differential operators*. SIAM.
- MEINSHAUSEN, N. & BÜHLMANN, P. (2006). High-dimensional graphs and variable selection with the lasso. *The annals of statistics* **34**, 1436–1462.
- MORRIS, J. S. (2015). Functional regression. *Annual Review of Statistics and Its Application* **2**, 321–359.
- RAMSAY, J. O. & SILVERMAN, B. W. (2005). *Functional data analysis*. Springer.

-
- RAMSAY, J. O., WICKHAM, H., GRAVES, S. & HOOKER, G. (2011). *Functional Data Analysis in R*. R package version 2.2.6.
- TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* **58**, 267–288.
- WANG, J.-L., CHIOU, J.-M. & MÜLLER, H.-G. (2016). Functional data analysis. *Annual Review of Statistics and its application* **3**, 257–295.
- YUAN, M. & LIN, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **68**, 49–67.
- ZOU, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical association* **101**, 1418–1429.

Ringraziamenti

Eccoci dunque alla fine di questo lavoro. La fine della tesi e la fine anche dell'Università. Si tratta quindi di un momento di transizione importante, dalla scuola al lavoro, in cui viene naturale fermarsi, riflettere su quanto fatto e quanto c'è da fare e ringraziare, perché ai traguardi così non ci si arriva da soli.

In questi anni ho studiato, parlato e discusso un bel po' di Statistica, grazie a docenti e compagni con i quali ho avuto la fortuna di confrontarmi quotidianamente. Un doveroso ringraziamento va dunque innanzitutto al professor Antonio Canale, per la sua disponibilità e i suoi consigli, per aver con pazienza ridimensionato i problemi a cui davo troppo peso e per aver creduto in me.

Ringrazio poi l'ARPAV per i dati forniti e il suo gentile personale di Teolo per le indicazioni e i preziosi suggerimenti.

Grazie a tutti gli amici di Statistica, che hanno riempito le mie giornate di questi ultimi anni con momenti di studio, lunghe chiacchierate post lezione, pranzi in giardino, *briefing* in aula studio, caffè da Aldon, pause alle macchinette, aperitivi a fine giornata e belle serate. Grazie quindi a Paolo il capo gita, Pietro, Gaspa, Giulia, Riccardo, Matteo, Alessio, Maurizio, Chiara, Anna e a tutti gli altri. In particolare, grazie ad Irene per aver condiviso con me quasi interamente questo periodo di tesi e per essere una delle più belle scoperte qui all'Università. Ringrazio Francesco per il suo ottimismo travolgente, la sua connaturata simpatia e la sua compagnia nelle domeniche patavine.

Dei ringraziamenti sinceri vanno poi a Chiarona, Gretix, Marix, Leonora, AmoElep e AmoElez, ossia alle ragazze che per tanto o per poco tempo sono state con me in casa di Terry e che hanno sopportato il mio caos e condiviso con

me la quotidianità . . . già mi mancate.

Grazie agli amici e alle amiche di sempre, con cui condivido tanti ricordi e vecchie esperienze, ma che tutt'ora sono tra le mie persone preferite in assoluto. Ringrazio Ellen, la mia roccia, per l'entusiasmo, la spontaneità, l'empatia e pure la sbadataggine. Serena la mia compagna di una vita, di cui mi ricordo da quando ho ricordi. E poi Alessandro, che meno male che è arrivato a Padova a condividere con me i budini della mensa e quei viaggi a Vienna tra velluti, specchi e mogano.

Infine un grazie immenso va alla mia famiglia. La mamma e il papà che sono i miei pilastri e i miei *fan* numero uno, che hanno sempre creduto in me più di quanto lo facessi io stessa. Grazie a Giulia, il mio modello, colei che più di tutti sa con pazienza ascoltare e capire le mie preoccupazioni e che mi ha regalato un nipotino meraviglioso. Grazie a nonna Renata e nonna Anna per la generosità e per tutti gli "stai serena" che mi sono stati amorevolmente ripetuti.

Sono quindi grata per tutte le persone, le esperienze e i luoghi che mi hanno accompagnata fino a qui e sono grata a me stessa per aver scelto questa facoltà. Dicono che la Statistica sia la risposta consapevole all'arroganza della certezza e forse proprio per questo la trovo una scienza così affascinante: con umiltà permette di mettere ordine, sbagliando ma con cognizione dei propri errori.

P.S. grazie anche ad Ada, che è stata qui sdraiata vicino a me ad ascoltarmi fino a che non ho scritto quest'ultima parola.