

Un modello di rete neurale profonda per la
stima della superficie di mortalità
dall'aspettativa di vita alla nascita

Indice

1	Modelli di valutazione per il longevity risk	5
1.1	Modellizzare l'incertezza presente nella dinamica della mortalità	6
1.2	Notazione	8
1.3	Modelli stocastici di tipo APC generalizzati	9
2	Il modello Lee-Carter e relative estensioni	12
2.1	Il modello Lee-Carter (1992)	12
2.1.1	Estensione del modello Lee-Carter: Brouhns et al. (2002)	15
2.1.2	Stima dei parametri	16
2.1.3	Proiezione	18
2.1.4	Approcci alternativi di stima	20
2.1.4.1	Second Stage Reestimation	20
2.1.4.2	Il metodo dei minimi quadrati pesati di Wilmoth	22
2.2	Il modello Cairns-Blake-Dowd (M5)	25
2.2.1	Stima dei parametri	25
2.2.2	Proiezione di $\kappa_t^{(1)}$ e $\kappa_t^{(2)}$	26
2.3	Il modello di Renshaw-Haberman (2006): Lee-Carter con effetto coorte	27
2.4	I modelli di Cairns-Blake-Dowd (M6-M7-M8) (2007)	28
2.5	Il modello Plat (2009)	30
3	Un modello di deep learning: le reti neurali	32
3.1	Modello della rete neurale	33
3.2	Caratteristiche generali	35
3.2.1	Tipi di architetture della rete	36
3.3	Back-propagation	42

<i>INDICE</i>	2
4 Stima della superficie di mortalità	45
4.1 Dati	46
4.2 Il modello	46
4.3 Risultati	49
4.3.1 Italia	49
4.3.2 USA	54
4.3.3 Giappone	59
4.3.4 Australia	63
4.3.5 Russia	68
4.3.6 Indici di bontà della stima	72
5 Conclusioni	75

Introduzione

L'aumento della longevità umana durante il ventesimo secolo ha suscitato un'attenzione sempre maggiore verso la modellizzazione e la proiezione dei tassi di mortalità e dell'aspettativa di vita, sia da parte dei demografi che degli attuari. Seguendo questa linea di ricerca, l'elaborato contribuisce alla letteratura proponendo un approccio innovativo in grado di stimare la superficie di mortalità a partire da una sola misura demografica riassuntiva. L'obiettivo è quello di descrivere ed analizzare un nuovo modello, basato su una rete neurale profonda (Deep Neural Network), capace di prevedere i tassi futuri di mortalità utilizzando come unico input l'aspettativa di vita alla nascita. In particolare, sono stati confrontati i tassi osservati con i risultati ottenuti attraverso ciascuno dei seguenti modelli:

- Il Lee-Carter, tra i più celebri in letteratura, che permette di stimare i tassi futuri utilizzando una matrice di tassi storici di mortalità, distinti per età e anno di calendario. Le analisi sono state effettuate utilizzando un'estensione del modello, proposta da Brouhns et al. (2002), in cui si ipotizza che i decessi siano distribuiti come una Poisson;
- La rete neurale profonda, ispirata alla struttura ed al funzionamento del sistema nervoso umano. La rete cattura come dati in input l'aspettativa di vita per ogni anno considerato, apprende la relazione nascosta nei dati e restituisce come output la superficie di mortalità richiesta per gli stessi anni.

In particolare, il lavoro è strutturato nel modo seguente: introdotto il concetto del longevity risk, viene analizzata la conseguente esigenza di modellizzare il fenomeno della mortalità, facendo particolare riferimento alla famiglia dei modelli generalized age-period-cohort (GAPC). Nel secondo capitolo si illustra il noto modello Lee-Carter, ripercorrendone le fasi di stima e di proiezione, partendo dalla sua versione originale (1992) e giungendo poi, allo

studio di tutte le relative estensioni volte a superarne i limiti. Nel terzo capitolo vengono introdotte ed approfondite le reti neurali, analizzandone le varie tipologie di struttura e le diverse caratteristiche. Nel quarto capitolo, viene proposto un nuovo modello basato su una rete neurale feedforward. Lo studio relativo alla capacità predittiva della rete è stato sperimentato utilizzando i dati forniti dallo Human Mortality Database relativi ad Italia, Usa, Giappone, Australia e Russia, facendo riferimento ad entrambi i sessi. Gli anni presi in esame sono quelli che vanno dal 1950 al 2014 per ciascun paese, ad eccezione della Russia per la quale non vi sono dati disponibili prima del 1959. I dati relativi agli anni che vanno dal 1950 al 2005 sono stati utilizzati per addestrare la rete, mediante la procedura di train-test. Colta la relazione nascosta tra i dati, la rete proietta i tassi futuri relativi agli anni che vanno dal 2006 al 2014, nella fase di validation. Infine, nel quinto capitolo, si riportano le conclusioni.

Capitolo 1

Modelli di valutazione per il longevity risk

Nel corso degli ultimi due secoli, i paesi sviluppati hanno registrato un persistente aumento dell'aspettativa di vita. Oeppen e Vaupel (2002) ad esempio, stimano un aumento di circa tre mesi l'anno dell'aspettativa di vita femminile alla nascita. Questo fenomeno, sebbene rappresenti da un lato un segno di progresso sociale, dall'altro costituisce una sfida per i governi, per i piani pensionistici privati e per gli assicuratori sulla vita, dato il suo impatto sui costi pensionistici e sanitari.

Attuari e demografi hanno presto riconosciuto i problemi causati dall'invecchiamento della popolazione e dall'aumento della longevità ed hanno quindi dedicato notevole attenzione allo sviluppo di tecniche statistiche per la modellizzazione e la proiezione dei tassi di mortalità.

Il profilo temporale della mortalità osservato negli ultimi decenni per ogni generica età x , evidenzia tre aspetti significativi: la rettangolarizzazione, l'espansione ed il più alto livello insieme, con una più larga dispersione della morte a giovani età (Pitacco (2004), Olivieri (2001)). La rettangolarizzazione (particolarmente marcata a partire dalle età adulte $x \geq 35$) consiste in una maggiore concentrazione della distribuzione di probabilità intorno al punto di Lexis (cioè alla moda della distribuzione), il quale a propria volta tende a coincidere con la durata massima di vita. L'espansione consiste invece nello spostamento in avanti del punto di Lexis che determina l'incertezza dell'ampiezza della rettangolarizzazione. Tali marcate tendenze si traducono in un allungamento dell'aspettativa di vita, fatta salva una forte volatilità dei tassi di mortalità nelle età giovanili.

Il longevity risk può essere definito ad un livello individuale o aggregato. Al livello individuale, si sostanzia nell'eventualità che la popolazione viva più a lungo di quanto previsto. Al livello aggregato, il longevity risk rappresenta invece, il rischio che l'insieme dei percettori di rendita vivano in media di più di quanto assunto nel pricing. I miglioramenti nei tassi di mortalità sono causati dall'effetto combinato dei due summenzionati fenomeni di rettangolarizzazione ed espansione.

1.1 Modellizzare l'incertezza presente nella dinamica della mortalità

Le fonti di rischio responsabili delle deviazioni osservabili ex-post tra i valori teorici forniti dal modello ed i valori osservati nella realtà, possono essere ricondotte a:

- Rischio di modello, determinato dall'inadeguatezza del metodo di proiezione a rappresentare l'effettivo trend di mortalità;
- Rischio di parametro, originato dall'incertezza presente nei valori dei parametri del modello di proiezione;
- Rischio di processo, generato dalla natura stocastica del processo (dei processi) temporale che governa (governano) l'evoluzione della mortalità;

La natura sistematica del longevity risk rende improponibile il ricorso a operazioni di diversificazione di portafoglio. Le tecniche sviluppate fino ad oggi per far fronte a tale rischio sono essenzialmente riconducibili a due approcci: la gestione del rischio in proprio o il trasferimento del rischio a soggetti terzi. In particolare, nel primo caso, le compagnie assicurative implementano tavole di mortalità proiettate.

La letteratura attuariale ha sviluppato numerosi approcci al fine di proiettare i tassi di mortalità; tra questi i metodi estrapolativi, basati sui trend storici della mortalità, costituiscono la classe di modelli di gran lunga più utilizzati dagli attuari. Essi si fondano sull'assunzione che l'evoluzione della mortalità osservata in passato si ripeterà in futuro. Nell'ambito dei metodi estrapolativi si distinguono quelli di tipo deterministico, che semplicemente estendono l'andamento della sopravvivenza passata al futuro, e quelli di tipo

stocastico, che operano previsioni sulla base di date distribuzioni di probabilità. Per ciò che concerne i modelli estrapolativi stocastici, i risultati delle proiezioni forniscono stime puntuali ed intervalli di confidenza per i tassi di mortalità futuri; essi permettono una valutazione di tipo stocastico.

Due famiglie di modelli molto noti in letteratura e maggiormente impiegati nelle valutazioni attuariali sono:

- Lee-Carter, 1992 e successive estensioni
- Cairns-Blake-Dowd, 2006 e successive estensioni

Uno degli approcci più influenti per la modellizzazione stocastica dei tassi di mortalità è il parsimonioso modello proposto da Lee e Carter (1992). Questo utilizza un'analisi a componenti principali per scomporre la matrice età-tempo dei tassi di mortalità in una combinazione bilineare dei parametri di età e periodo, i quali vengono trattati come serie temporali per produrre proiezioni. Il modello Lee-Carter ha ispirato numerose varianti ed estensioni. Ad esempio, Lee e Miller (2001), Booth et al. (2002) e Brouhns et al. (2002) hanno proposto approcci di stima alternativi al fine di migliorare la bontà di adattamento e la proprietà di previsione del modello. In dettaglio, Brouhns et al. (2002) propongono un approccio statistico più formale per la stima dei parametri, incorporando il modello Lee-Carter in un'impostazione di regressione di tipo Poisson. Altri autori hanno esteso il modello Lee-Carter includendo termini aggiuntivi, come ad esempio componenti multiple bilineari dei parametri di età e periodo (Renshaw e Haberman 2003; Hyndman and Ullah 2007), oppure termini con effetto coorte (Renshaw and Haberman 2006).

Il modello Cairns-Blake-Dowd (CBD) a due fattori, introdotto da Cairns et al. (2006), è una delle varianti più importanti del modello Lee-Carter. Il modello CBD si basa sulla linearità del logit delle probabilità di morte di un anno alle età più anziane. In particolare esso assume che, per un dato anno, il logit della probabilità di morte di un anno sia una funzione lineare dell'età, e considera i parametri di intercetta e pendenza, attraverso gli anni, dei processi stocastici. Cairns et al. (2009) introducono tre estensioni al modello originale CBD incorporando combinazioni di un termine di età quadratico e un termine di effetto di coorte. Plat (2009) ha combinato le caratteristiche dei modelli CBD e Lee-Carter per produrre un modello che risulti adatto per intere fasce di età e che catturi l'effetto coorte.

Data l'abbondanza e il rapido aumento del numero di modelli stocastici di mortalità proposti nella letteratura, si è tentato di trovare dei punti in comune tra questi. Hunt and Blake (2015) analizzano la struttura dei modelli di mortalità e descrivono una struttura di tipo age-period-cohort che comprende la stragrande maggioranza dei metodi stocastici di mortalità. Currie (2016) mostra come molti modelli di mortalità possano essere espressi in termini di modelli lineari generalizzati o modelli non lineari generalizzati.

1.2 Notazione

Si indichi con D_{xt} la variabile aleatoria rappresentante il numero di decessi in una popolazione all'età x durante l'anno di calendario t . Sia d_{xt} invece, il numero osservato di decessi, E_{xt}^c l'esposizione centrale al rischio all'età x nell'anno t , e E_{xt}^0 la corrispondente esposizione iniziale al rischio. La probabilità di morte di un anno per un individuo di età x nell'anno di calendario t , chiamata q_{xt} , può essere stimata come

$$\hat{q}_{xt} = \frac{d_{xt}}{E_{xt}^0}.$$

La forza di mortalità ed il tasso centrale di mortalità sono indicati rispettivamente con μ_{xt} e m_{xt} ; la stima empirica di quest'ultimo è

$$\hat{m}_{xt} = \frac{d_{xt}}{E_{xt}^c}.$$

Ipotizzando che la forza di mortalità sia costante per ogni anno di età e di calendario, ovvero dall'età x all'età $x + 1$ e dall'anno t all'anno $t + 1$, la forza di mortalità μ_{xt} e il tasso di mortalità m_{xt} coincidono.

Nel pacchetto R StMoMo e in seguito, si assume che i decessi, d_{xt} , le esposizioni centrali, E_{xt}^c , o le esposizioni iniziali, E_{xt}^0 , siano disponibili in un formato di allineamento rettangolare comprendente le età (sulle righe) $x = x_1, x_2, \dots, x_k$ e gli anni di calendario (sulle colonne) $t = t_1, t_2, \dots, t_n$. Quando sono disponibili solo le esposizioni centrali e sono richieste le esposizioni iniziali (o viceversa), è possibile approssimare le esposizioni iniziali aggiungendo la metà del numero dei decessi riferito alle esposizioni centrali, cioè

$$E_{xt}^0 \approx E_{xt}^c + \frac{1}{2}d_{xt}.$$

Quando il contesto è chiaro, si può scrivere E_{xt} per riferirsi a E_{xt}^c o a E_{xt}^0 .

1.3 Modelli stocastici di tipo APC generalizzati

Alcuni autori hanno recentemente cercato di identificare le somiglianze tra i vari modelli statistici proposti. Ad esempio, Hunt e Blake (2015) descrivono una struttura di tipo age-period-cohort (APC) che comprende la stragrande maggioranza dei modelli di mortalità stocastica. In un altro interessante contributo, Currie (2016) mostra che molti modelli comuni di mortalità possono essere espressi nella terminologia standard dei modelli lineari o non lineari generalizzati. In questa sezione viene definita la famiglia dei modelli stocastici di mortalità per età-coorte generalizzati (GAPC).

Analogamente ai modelli lineari generalizzati (cfr., ad esempio, McCullagh e Nelder (1989)), un modello di mortalità stocastica GAPC è composto da quattro componenti:

1. La *componente casuale*: i decessi D_{xt} sono distribuiti secondo una Poisson o una binomiale, in modo tale che:

$$D_{xt} \sim \text{Poisson}(E_{xt}^c, \mu_{xt})$$

oppure

$$D_{xt} \sim \text{Binomial}(E_{xt}^0, q_{xt})$$

con

$$\mathbb{E}(D_{xt}/E_{xt}^c) = \mu_{xt}$$

e

$$\mathbb{E}(D_{xt}/E_{xt}^0) = q_{xt}.$$

2. La *componente sistematica*: gli effetti dell'età x , dell'anno di calendario t e dell'anno di nascita (coorte) $c = t - x$ sono catturati attraverso un predittore η_{xt} , dato da:

$$\eta_{xt} = \alpha_x + \sum_{i=1}^N \beta_x^{(i)} \kappa_t^{(i)} + \beta_x^{(0)} \gamma_{t-x}.$$

Dove:

- il termine α_x descrive il comportamento della mortalità al variare dell'età.

- $N \geq 0$ è un numero intero che indica il numero di termini age-period che descrivono l'andamento della mortalità. L'indice temporale $\kappa_t^{(i)}$, $i = 1, \dots, N$, spiega la variazione della mortalità nel tempo e $\beta_x^{(i)}$ descrive per ogni età come la mortalità reagisce al variare del parametro temporale $\kappa_t^{(i)}$.
- Il parametro γ_{t-x} rappresenta l'effetto coorte e $\beta_x^{(0)}$ descrive per ogni età come la mortalità reagisce al variare dell'effetto coorte.

I parametri $\beta_x^{(i)}$, $i = 0, 1, \dots, N$, possono essere funzioni predefinite dell'età, ad esempio $\beta_x^{(i)} \equiv f^i(x)$, come nei modelli di tipo CBD, oppure possono essere termini non parametrici senza alcuna struttura che deve essere stimata, come nel modello Lee-Carter. Nella famiglia dei modelli di tipo GAPC si assume che gli indici periodali $\kappa_t^{(i)}$ $i = 1, \dots, N$ e l'indice di coorte γ_{t-x} siano dei processi stocastici. Questa è la caratteristica chiave che consente la proiezione stocastica dei modelli GAPC e quindi la generazione di previsioni probabilistiche dei tassi futuri di mortalità.

3. La *link function* g che associa la componente casuale a quella sistematica, tale che:

$$g\left(\mathbb{E}\left(\frac{D_{xt}}{E_{xt}}\right)\right) = \eta_{xt}.$$

Sebbene siano possibili diverse link functions, risulta conveniente utilizzare la cosiddetta canonical link ed associare la distribuzione di Poisson con la log link function e quella Binomiale con la logit link function.

4. L'*insieme di vincoli sui parametri*: la maggior parte dei modelli stocastici necessitano di vincoli sui parametri per poter essere calibrati, altrimenti pongono problemi di identificabilità dei parametri. Questi vincoli di parametro vengono applicati attraverso una funzione di vincolo v , la quale associa un vettore arbitrario di parametri

$$\theta := \left(\alpha_x, \beta_x^{(1)}, \dots, \beta_x^{(N)}, \kappa_t^{(1)}, \dots, \kappa_t^{(N)}, \beta_x^{(0)}, \gamma_{t-x}\right)$$

ad un vettore di parametri trasformati

$$v(\theta) = \tilde{\theta} = \left(\tilde{\alpha}_x, \tilde{\beta}_x^{(1)}, \dots, \tilde{\beta}_x^{(N)}, \tilde{\kappa}_t^{(1)}, \dots, \tilde{\kappa}_t^{(N)}, \tilde{\beta}_x^{(0)}, \tilde{\gamma}_{t-x}\right)$$

soddisfacendo i vincoli del modello senza alcun effetto sul predittore η_{xt} : θ e $\tilde{\theta}$ producono lo stesso η_{xt} .

*CAPITOLO 1. MODELLI DI VALUTAZIONE PER IL LONGEVITY RISK*11

La maggior parte dei modelli stocastici di mortalità proposti in letteratura appartengono alla famiglia GAPC. Tra questi si segnala il modello Lee-Carter originale, le relative estensioni proposte in Renshaw e Haberman (2003, 2006), il modello CBD originale e le estensioni di Cairns et al. (2009). Inoltre, tutte le strutture di modello considerate in Haberman e Renshaw (2011), Lovász (2011) e van Berkum et al. (2014), così come i modelli di Plat (2009), Aro e Pennanen (2011), O'Hare e Li (2012), Borger et al. (2013) e Alai e Sherris (2014), possono essere ricondotte alla classe dei modelli di tipo GAPC.

Capitolo 2

Il modello Lee-Carter e relative estensioni

Il metodo proposto da Lee e Carter è considerato il «principale modello statistico di mortalità [in ottica previsionale] nella letteratura demografica» (Deaton e Paxson, 2004). La prima versione del modello è stata sviluppata dagli autori nel 1992 utilizzando i dati sulla mortalità degli Stati Uniti nel periodo 1933-1987. Tuttavia il metodo viene ora applicato, da molti paesi e per ogni genere di causa, ai dati sulla mortalità, ben oltre l'applicazione per cui era stato progettato.

Nel corso degli anni, sono stati proposti diversi miglioramenti di questo modello, ma la versione originale, insieme all'estensione fornita da Brouhns et al. (2002), è ancora considerata il punto di riferimento per il confronto con gli sviluppi futuri.

2.1 Il modello Lee-Carter (1992)

Il modello ipotizza un andamento esponenziale del tasso centrale di mortalità per ogni età. Quest'ultimo può essere interpretato come una probabilità di decesso centrale, cioè calcolata alla metà del periodo in considerazione (mediante il rapporto tra decessi e numero medio di persone). L'andamento esponenziale del tasso centrale di mortalità ricalca perfettamente il fenomeno della decelerazione: la vita attesa per un soggetto di età x cresce nel tempo ma la velocità con cui ciò avviene tende a diminuire.

Si indichi con $\ln m_{xt}$ il logaritmo del tasso centrale di mortalità di un paese, calcolato considerando il gruppo di età x ($x = 1, \dots, X$) al tempo t ($t = 1, \dots, T$).

Nella sua versione originale, in cui i tassi centrali di mortalità hanno una forma log-bilineare, il modello è caratterizzato dalla seguente relazione (Lee e Carter, 1992):

$$(1) \quad \ln m_{xt} = \alpha_x + \beta_x \kappa_t + \epsilon_{xt}$$

ove:

- α_x descrive il comportamento della mortalità al variare dell'età
- β_x descrive per ogni età come la mortalità reagisce al variare di κ_t
- κ_t indice della variazione della mortalità nel tempo
- ϵ_{xt} termine di errore: errori indipendenti ed identicamente distribuiti con distribuzione $N(0, \sigma_\epsilon^2)$

α_x, β_x e κ_t sono parametri da stimare, ϵ_{xt} è un insieme di termini di disturbo casuali.

La seguente parametrizzazione tuttavia non è unica, poichè è invariante rispetto alle trasformazioni:

$$\beta_x \rightsquigarrow c\beta_x \quad \kappa_t \rightsquigarrow \frac{1}{c}\kappa_t \quad \forall c \in \mathbb{R}, c \neq 0$$

$$\alpha_x \rightsquigarrow \alpha_x - \beta_x c \quad \kappa_t \rightsquigarrow \kappa_t + c \quad \forall c \in \mathbb{R}.$$

Questo non costituisce tuttavia un ostacolo concettuale; significa semplicemente che la probabilità associata al modello ha un numero infinito di massimi equivalenti, ciascuno dei quali produrrebbe previsioni identiche. Nella pratica, si sceglie una parametrizzazione arbitraria, ma coerente, sufficiente per l'identificazione. Il modello ha dunque bisogno dei seguenti due vincoli sui parametri per poter essere calibrato, altrimenti pone problemi di identificabilità dei parametri stessi:

$$\sum_x \beta_x = 1$$

e

$$\sum_t \kappa_t = 0$$

I parametri κ_t stimati sono poi modellizzati e proiettati come una serie temporale stocastica utilizzando i modelli ARIMA.

Semplificando la notazione, si consideri m_{xt} direttamente il logaritmo del tasso centrale di mortalità di un paese.

Il vincolo $\sum_t \kappa_t = 0$ mostra che il parametro α_x rappresenta semplicemente una media empirica di m_{xt} , lungo l'intero periodo di osservazione, $\alpha_x = \bar{m}_x$.

Pertanto è possibile riscrivere il modello in termini di media centrata del tasso di log-mortalità,

$$\tilde{m}_{xt} = m_{xt} - \bar{m}_x.$$

Il modello Lee-Carter implica che i termini di disturbo ϵ_{xt} siano normalmente distribuiti; per tale ragione, è possibile riscrivere l'equazione (1) come modello moltiplicativo per il profilo di età centrato:

$$(2) \quad \tilde{m} \sim \mathcal{N}(\bar{\mu}_{xt}, \sigma^2)$$

$$\mathbb{E}(\tilde{m}_{xt}) \equiv \bar{\mu}_{xt} = \beta_x \kappa_t.$$

In questa espressione, vengono utilizzati solo $X + T$ parametri ($\beta_x \kappa_t$, per ciascun x e t , rappresentati sui margini in basso e a destra della matrice \tilde{m}) per approssimare gli elementi della seguente matrice $X \times T$:

	1990	1991	1992	1993	1994	
5	$\tilde{m}_{5,0}$	$\tilde{m}_{5,1}$	$\tilde{m}_{5,2}$	$\tilde{m}_{5,3}$	$\tilde{m}_{5,4}$	β_5
10	$\tilde{m}_{10,0}$	$\tilde{m}_{10,1}$	$\tilde{m}_{10,2}$	$\tilde{m}_{10,3}$	$\tilde{m}_{10,4}$	β_{10}
15	$\tilde{m}_{15,0}$	$\tilde{m}_{15,1}$	$\tilde{m}_{15,2}$	$\tilde{m}_{15,3}$	$\tilde{m}_{15,4}$	β_{15}
20	$\tilde{m}_{20,0}$	$\tilde{m}_{20,1}$	$\tilde{m}_{20,2}$	$\tilde{m}_{20,3}$	$\tilde{m}_{20,4}$	β_{20}
25	$\tilde{m}_{25,0}$	$\tilde{m}_{25,1}$	$\tilde{m}_{25,2}$	$\tilde{m}_{25,3}$	$\tilde{m}_{25,4}$	β_{25}
30	$\tilde{m}_{30,0}$	$\tilde{m}_{30,1}$	$\tilde{m}_{30,2}$	$\tilde{m}_{30,3}$	$\tilde{m}_{30,4}$	β_{30}
35	$\tilde{m}_{35,0}$	$\tilde{m}_{35,1}$	$\tilde{m}_{35,2}$	$\tilde{m}_{35,3}$	$\tilde{m}_{35,4}$	β_{35}
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
80	$\tilde{m}_{80,0}$	$\tilde{m}_{80,1}$	$\tilde{m}_{80,2}$	$\tilde{m}_{80,3}$	$\tilde{m}_{80,4}$	β_{80}
	κ_0	κ_1	κ_2	κ_3	κ_4	

Il modello Lee-Carter approssima $\tilde{m}_{5,0}$, ad esempio, con il prodotto dei parametri alla fine della prima riga e della prima colonna $\beta_5\kappa_0$.

Pertanto il modello Lee-Carter può anche essere considerato un caso particolare dei modelli log-lineari per tabelle di contingenza (Bishop, Fienberg e Holland, 1975; King, 1989: cap. 6), dove gli elementi della matrice sono approssimati con le stime dei parametri dalle marginali. Si tratta effettivamente della versione più semplice dei modelli per tabelle di contingenza, nei quali si assume l'indipendenza di righe (fasce di età) e colonne (periodi di tempo), ed in cui il generico valore atteso della matrice è semplicemente dato dal prodotto dei parametri ricavati dalle rispettive marginali:

$$\mathbb{E}(\tilde{m}_{xt}) = \beta_x \kappa_t.$$

Questa ipotesi tuttavia, risulterebbe appropriata se la variabile rappresentata sulle righe della matrice fosse indipendente da quella rappresentata sulle colonne. Per il tasso di log-mortalità ciò significa assumere l'assenza di interazioni tra l'età ed il tempo, e cioè ipotizzare che i parametri β_x e κ_t rimangano fissi nel tempo per ogni x e per ogni t , rispettivamente.

Si tratta di un'assunzione poco realistica soprattutto per età elevate, dove è presente una maggiore variabilità della mortalità a causa dell'esiguo numero di decessi. Per tale ragione vengono proposte varie estensioni del modello Lee-Carter, che consentono oltre che dei cambiamenti strutturali, variazioni di β_x nel tempo.

2.1.1 Estensione del modello Lee-Carter: Brouhns et al. (2002)

Il modello Lee-Carter assume implicitamente che gli errori casuali ϵ_{xt} siano omoschedastici (stessa varianza rispetto all'età); si tratta tuttavia di un'ipotesi poco realistica per età elevate, dove è presente una maggiore variabilità della mortalità a causa dell'esiguo numero di decessi.

La proposta di Brouhns et al. (2002) consiste nell'implementare il modello Lee-Carter ipotizzando che i decessi siano distribuiti come una Poisson:

$$D_{xt} \sim \text{Poisson}(E_{xt}, m_{xt}).$$

Rispetto al Lee-Carter originario, il termine di errore additivo ϵ_{xt} viene sostituito da una variazione casuale del numero di decessi di tipo Poisson. Ciò rappresenta un'ipotesi realistica per età elevate.

2.1.2 Stima dei parametri

I parametri β_x e κ_t del modello, possono essere stimati con il metodo della massima verosimiglianza.

Tuttavia, i vincoli imposti sui parametri possono causare un malfunzionamento dei programmi di ottimizzazione standard.

Fortunatamente, come sottolineato da Lee e Carter, la soluzione ottimale del problema può essere facilmente trovata tramite la scomposizione in componenti principali (Singular Value Decomposition) della matrice centrata dei profili di età, $\tilde{m} = BLU'$, dove la prima colonna di B è costituita dai valori stimati di β_x , mentre le stime di κ_t sono $\beta' \tilde{m}_t$. Se la decomposizione SVD di \tilde{m} non è disponibile, si può calcolare come autovettore normalizzato della matrice $C \equiv \tilde{m} \tilde{m}'$ corrispondente all'autovalore più grande.

La giustificazione teorica sottostante la procedura, sia se si ricorre alla scomposizione in componenti principali (SVD) che agli autovalori, rimane il metodo della massima verosimiglianza.

Nella pratica, Lee e Carter suggeriscono, dopo aver stimato i parametri β e κ , un'ulteriore fase di stima per i parametri κ_t utilizzando un diverso criterio.

Questo passaggio di rivalutazione, spesso chiamato "second stage estimation", non sempre fornisce un'unica soluzione per il criterio delineato in Lee and Carter (1992). Recentemente sono stati proposti diversi criteri (Lee e Miller, 2001; Wilmoth, 1993); alcuni ricercatori decidono tuttavia di saltare del tutto questa seconda fase.

Più nel dettaglio, la stima dei parametri del Lee-Carter prevede:

- Definizione di una funzione obiettivo da minimizzare (metodo dei minimi quadrati - Least Squares):

$$\vartheta_{LS}(\alpha, \beta, \kappa) = \sum_{x=x_1}^{x_m} \sum_{t=t_1}^{t_n} (\ln \hat{m}_x(t) - \alpha_x - \beta_x \kappa_t)^2,$$

dove gli $\hat{m}_x(t)$ indicano i tassi osservati.

- Scomposizione in componenti principali (singular value decomposition):

Si pone pari a 0 la derivata $\frac{\partial}{\partial \alpha_x} \vartheta_{LS}$ ovvero:

$$\sum_{t=t_1}^{t_n} \ln \hat{m}_x(t) = (t_n - t_1 + 1) \alpha_x + \beta_x \sum_{t=t_1}^{t_n} \kappa_t.$$

Per effetto del vincolo $\sum_t \kappa_t = 0$ si ottiene:

$$\hat{\alpha}_x = \frac{1}{t_n - t_1 + 1} \sum_{t=t_1}^{t_n} \ln \hat{m}_x(t).$$

I parametri β e κ si ottengono effettuando la scomposizione in componenti principali del primo termine della matrice $\ln \hat{m}_x(t) - \hat{\alpha}_x$.

Si considera la matrice dei tassi centrali di mortalità:

$$\mathbf{M} = \begin{pmatrix} m_{x_1}(t_1) & \cdots & m_{x_1}(t_n) \\ \vdots & \ddots & \vdots \\ m_{x_m}(t_1) & \cdots & m_{x_m}(t_n) \end{pmatrix}$$

e si definisce la matrice da minimizzare:

$$\mathbf{Z} = \ln \mathbf{M} - \boldsymbol{\alpha} = \begin{pmatrix} \ln \hat{m}_{x_1}(t_1) - \hat{\alpha}_{x_1} & \cdots & \ln \hat{m}_{x_1}(t_n) - \hat{\alpha}_{x_1} \\ \vdots & \ddots & \vdots \\ \ln \hat{m}_{x_m}(t_1) - \hat{\alpha}_{x_m} & \cdots & \ln \hat{m}_{x_m}(t_n) - \hat{\alpha}_{x_m} \end{pmatrix}$$

I parametri β_x e κ_t sono tali da minimizzare la funzione:

$$\vartheta_{LS}(\boldsymbol{\beta}, \boldsymbol{\kappa}) = \sum_{x=x_1}^{x_m} \sum_{t=t_1}^{t_n} (z_{xt} - \beta_x \kappa_t)^2.$$

Occorre infine trovare gli autovettori e gli autovalori della matrice \mathbf{Z} .

- Aggiustamento dei parametri κ_t ricalibrati sulla distribuzione per età dei decessi osservati.

Si ipotizza che i decessi siano distribuiti secondo una Poisson:

$$D_{xt} \sim \text{Poisson}(E_{xt}, m_{xt})$$

I parametri κ_t devono quindi essere tali da massimizzare la Poisson log-verosimiglianza rispetto a ζ :

$$\sum_{x=x_1}^{x_m} \left(D_{xt} \ln \left(E_{xt} \exp \left(\hat{\alpha}_x + \hat{\beta}_x \zeta \right) \right) - E_{xt} \exp \left(\hat{\alpha}_x + \hat{\beta}_x \zeta \right) \right)$$

oppure da minimizzare la devianza:

$$\mathcal{D} = 2 \sum_{x=x_1}^{x_m} \left(D_{xt} \ln \frac{D_{xt}}{\hat{D}_{xt}} - \left(D_{xt} - \hat{D}_{xt} \right) \right)$$

dove D_{xt} indica i decessi osservati e $\hat{D}_{xt} = E_{xt} \exp \left(\hat{\alpha}_x + \hat{\beta}_x \zeta \right)$.

2.1.3 Proiezione

Per produrre previsioni sulla mortalità, Lee e Carter ipotizzano che i parametri β_x rimangano costanti nel tempo ed utilizzano le previsioni di $\hat{\kappa}_t$ da un modello di serie temporali univariate. Dopo aver testato diverse specifiche del modello ARIMA, il modello più appropriato per i loro dati, risulta essere un *random walk with drift*.

Nonostante possano essere preferibili altri modelli ARIMA per diversi dati, per κ_t si sceglie quasi esclusivamente la *random walk with drift*. Si considera il seguente modello:

$$\hat{\kappa}_t = \hat{\kappa}_{t-1} + \theta + \xi_t$$

$$\xi_t \sim \mathcal{N}(0, \sigma_{rw}^2)$$

dove θ è noto come il parametro di deriva (drift) e la rispettiva stima di massima verosimiglianza è semplicemente

$$\hat{\theta} = \frac{(\hat{\kappa}_T - \hat{\kappa}_1)}{(T - 1)},$$

la quale dipende unicamente dalla prima ed ultima stima di κ . Per prevedere due periodi in anticipo, si inserisce la stima del parametro di deriva $\hat{\theta}$ e si sostituisce la definizione di $\hat{\kappa}_{t-1}$ spostato indietro nel tempo di un periodo:

$$\begin{aligned} \hat{\kappa}_t &= \hat{\kappa}_{t-1} + \hat{\theta} + \xi_t \\ &= \left(\hat{\kappa}_{t-2} + \hat{\theta} + \xi_{t-1} \right) + \hat{\theta} + \xi_t \\ &= \hat{\kappa}_{t-2} + 2\hat{\theta} + (\xi_{t-1} + \xi_t) \end{aligned}$$

Per prevedere $\hat{\kappa}_t$ al tempo $T + (\Delta t)$, con i dati disponibili fino al periodo T , si esegue la stessa procedura iterativamente (Δt) volte e si ottiene

$$\begin{aligned} \hat{\kappa}_{T+(\Delta t)} &= \hat{\kappa}_T + (\Delta t) \hat{\theta} + \sum_{l=1}^{(\Delta t)} \xi_{T+l-1} \\ &= \hat{\kappa}_T + (\Delta t) \hat{\theta} + \sqrt{(\Delta t)} \xi_t, \end{aligned}$$

dove la seconda riga è una semplificazione resa possibile dal fatto che le variabili casuali ξ_t in questo modello, sono assunte indipendenti e con stessa varianza. La seconda uguaglianza indica che gli errori standard per la previsione aumentano con la radice quadrata della distanza dall'orizzonte di previsione (Δt) . Questi sono gli errori standard condizionali, i quali risulterebbero maggiori se venisse inclusa anche l'incertezza della stima.

A partire da questo modello, è possibile ottenere le stime puntuali di previsione, le quali seguono una linea retta come funzione di (Δt) , con pendenza $\hat{\theta}$:

$$\mathbb{E} [\hat{\kappa}_{T+(\Delta t)} | \hat{\kappa}_1, \dots, \hat{\kappa}_T] \equiv \mu_{T+(\Delta t)} = \hat{\kappa}_T + (\Delta t) \hat{\theta}$$

Il modello Lee-Carter per i κ è dunque molto semplice: si estrapola da una linea retta tracciata attraverso il primo $\hat{\kappa}_1$ e l'ultimo $\hat{\kappa}_T$. Tutti gli altri $\hat{\kappa}$ vengono di conseguenza ignorati.

Tali espressioni vengono ora inserite nell'equazione $\mathbb{E}(\tilde{m}_{xt}) \equiv \bar{\mu}_{xt} = \beta_x \kappa_t$, in modo tale da fornire una previsione di stima puntuale per la log-mortalità:

$$\begin{aligned} \mu_{t+(\Delta T)} &= \bar{m} + \hat{\beta} \hat{\kappa}_{T+(\Delta t)} \\ &= \bar{m} + \hat{\beta} \left[\hat{\kappa}_T + (\Delta t) \hat{\theta} \right]. \end{aligned}$$

Ad esempio, dati i tassi osservati dal 1950 al 2000, il modello Lee-Carter calcola la previsione per l'anno 2030 come segue:

$$\begin{aligned} \hat{\mu}_{2030} &= \bar{m} + \hat{\beta} \times \left[\hat{\kappa}_{2000} + 30\hat{\theta} \right] \\ &= \bar{m} + \hat{\beta} \times \left[\hat{\kappa}_{2000} + 30 \frac{(\hat{\kappa}_{2000} - \hat{\kappa}_{1950})}{50} \right]. \end{aligned}$$

2.1.4 Approcci alternativi di stima

E' possibile stimare i parametri del modello Lee-Carter utilizzando degli approcci alternativi.

2.1.4.1 Second Stage Reestimation

Nell' articolo originale, Lee e Carter (1992), suggeriscono un metodo alternativo di stima per l'indice di mortalità κ_t . Dopo aver descritto il criterio, vengono evidenziati alcuni problemi derivanti dallo stesso approccio, non riconosciuti in precedenza. Gli autori notano che una volta stimati β_x e κ_t , il numero osservato di decessi totale $d_t \equiv \sum_x d_{xt}$ non garantisce l'equivalenza col numero totale di decessi \hat{d}_t previsti dal modello:

$$\hat{d}_t = \sum_x p_{xt} e^{\bar{m}_x + \hat{\beta}_x \hat{\kappa}_t}.$$

Di conseguenza, è stato proposto di calcolare una nuova stima κ_t , per ogni anno t , mediante la ricerca di quel valore che rende il numero osservato

di decessi pari a quello previsto. Definendo questa nuova stima di κ_t come $\hat{\kappa}_t^*$, Lee e Carter suggeriscono di impostare:

$$d_t = \sum_x p_{xt} e^{\bar{m}_x + \hat{\beta}_x \hat{\kappa}_t^*}.$$

Questo modo di stimare κ_t presenta numerosi vantaggi, che sono descritti nel dettaglio in Lee e Carter (1992). Questi possono essere utili nella rappresentazione della tabella di vita dei dati, in particolare per i casi in cui sono noti per alcuni anni solo i tassi di mortalità totali, piuttosto che quelli specifici per età.

Un problema con la stima di κ_t , data dall'equazione di d_t , è che quest'ultima può avere 0, 1 o 2 soluzioni per κ_t^* , a seconda dei valori di $\hat{\beta}$. Quando ha un'unica soluzione non vi sono problemi; quando ne fornisce 0 o 2, l'approccio tuttavia, può risultare logicamente incoerente ed empiricamente inutile.

Un numero di soluzioni diverse da 1 si presentano solo quando i valori stimati $\hat{\beta}_x$ non hanno tutti lo stesso segno. Un segno non uniforme per i valori di $\hat{\beta}_x$ indica un aumento della mortalità in alcune fasce di età e una diminuzione in altre. Ciò è raro quando si predice la mortalità per tutte le cause, che è diminuita più o meno in tutto il mondo, con poche eccezioni come l'Ungheria. Pertanto, se il modello viene utilizzato per la mortalità negli Stati Uniti, per cui è stato progettato il Lee-Carter, questo problema si presenta solo occasionalmente. Nella mortalità per causa specifica, i segni non uniformi in $\hat{\beta}_x$ sono molto più comuni e quindi la rivalutazione sarebbe sconsigliata. Per vedere da dove deriva questo risultato, vengono partizionati i coefficienti $\hat{\beta}_x$ in due gruppi, quelli con segno positivo, che denotiamo

$$\hat{\beta}_x^+ = | \hat{\beta}_x^+ |,$$

e quelli con segno negativo

$$\hat{\beta}_x^- = - | \hat{\beta}_x^- |.$$

Si riscriva l'equazione d_t come segue:

$$d_t = \sum_x p_{xt} e^{\bar{m}_x} e^{\hat{\beta}_x^+ \hat{\kappa}_t^*} + \sum_x p_{xt} e^{\bar{m}_x} e^{\hat{\beta}_x^- \hat{\kappa}_t^*}.$$

Quando tutti i $\hat{\beta}_x$ hanno lo stesso segno, a destra dell'equazione si ha una somma di esponenziali in $\hat{\kappa}_t^*$, e quindi una funzione monotona di $\hat{\kappa}_t^*$, con intervallo $(0, +\infty)$. Ciò garantisce l'unicità della soluzione.

Quando i $\hat{\beta}_x$ hanno segni diversi, a destra dell'equazione si ha la somma degli esponenziali, alcuni in aumento ed altri in calo. Come è noto, tale funzione è a forma di U, avente come minimo un numero strettamente positivo. Pertanto, se d_t è abbastanza grande, una linea orizzontale all'altezza d_t intersecherà la funzione a forma di U in due punti: si ottengono così due soluzioni. Se d_t invece è abbastanza piccolo, passerà completamente al di sotto della funzione a forma di U, e il problema non avrà alcuna soluzione.

Una strategia alternativa è dovuta a Lee e Miller (2001), i quali propongono per κ_t una seconda fase di stima, in modo che il valore previsto e osservato dell'aspettativa di vita alla nascita, piuttosto che il numero totale di decessi, si equivalgano ogni anno.

Questo approccio non può essere applicato alla mortalità specifica per causa. Fortunatamente però, la rivalutazione non è una caratteristica cruciale del modello Lee-Carter.

2.1.4.2 Il metodo dei minimi quadrati pesati di Wilmoth

Alternativamente alla procedura di stima di base proposta da Lee-Carter, in Wilmoth (1993) vengono descritti due ulteriori approcci.

L'intuizione di Wilmoth consiste nel motivare la differenza tra il numero di decessi ottenuto dal modello e il numero effettivamente osservato: tale fenomeno si manifesta in quanto le stime di κ_t vengono calcolate minimizzando l'errore dei minimi quadrati sulla mortalità logaritmica, piuttosto che sulla mortalità. Di conseguenza, gruppi di età con piccoli numeri di decessi, ricevono lo stesso peso delle fasce d'età con un numero elevato, anche se il loro contributo sul totale è molto piccolo.

Quindi, piuttosto che compiere un'ulteriore fase di stima per κ_t , come descritto in Lee e Carter (1992), Wilmoth propone due strategie alternative: una tecnica ponderata del minimo quadrato (WLS) e una di massima verosimiglianza (MLE). Entrambi gli approcci hanno il vantaggio, rispetto alla formulazione originale Lee-Carter, di affrontare naturalmente il caso in cui il numero osservato di decessi è zero, il quale si verifica quando si analizzano dati specifici per causa e / o quando si ha a che fare con piccoli paesi. Mentre la tecnica MLE è statisticamente valida, è possibile dimostrare che per la procedura WLS non vale lo stesso.

Alla base di tale approccio vi è la consapevolezza di poter ponderare il primo step del Lee-Carter in modo tale da avvicinare il più possibile i decessi osservati a quelli previsti. Nello specifico, Wilmoth suggerisce di trovare i

parametri α_x, β_x e κ_t del modello Lee-Carter come soluzione del problema dei minimi quadrati ponderati (WLS)

$$\min_{\alpha_x, \beta_x, \kappa_t} \sum_{xt} d_{xt} (m_{xt} - \alpha_x - \kappa_t \beta_x)^2,$$

dove d_{xt} è il numero osservato di decessi nella fascia di età x e al tempo t , e dove i parametri κ_t e β_x soddisfano i soliti vincoli imposti dal modello Lee-Carter.

Valutando l'errore per il numero osservato di decessi, l'espressione di sopra attribuisce maggior peso a gruppi di età e di anni con un gran numero di morti, e le stime risultanti hanno maggiori probabilità di adattarsi al numero totale di decessi ogni anno.

E' interessante notare altre due caratteristiche di questa procedura. In primo luogo, con questo approccio viene meno il problema per cui la log-mortalità non può essere definita quando il numero di decessi è zero. Sebbene per $d_{xt} = 0$ l'espressione $m_{xt} = \ln(d_{xt}/p_{xt})$ non abbia senso, le stesse osservazioni vengono scartate dal problema di minimo poiché ricevono un peso nullo. Una seconda caratteristica interessante è data dalla semplicità con cui è possibile scrivere le condizioni del primo ordine corrispondenti (Wilmoth, 1993: 3).

Sfortunatamente, tale procedura non risulta essere statisticamente valida e questi vantaggi possono essere superati dalla distorsione indotta dallo stimatore. In particolare, Wilmoth interpreta il problema di minimo come la log-verosimiglianza di un modello con disturbi distribuiti come una normale, eteroschedastici e con varianza proporzionale a $\frac{1}{d_{xt}}$. Questa affermazione è tuttavia errata, in quanto la varianza dei disturbi non può essere (inversamente) proporzionale al numero osservato di decessi. Una probabilità ponderata valida, deve usare pesi esogeni, ma il numero di decessi è ovviamente una variabile casuale. Come tale, le stime risultanti da questo problema di minimizzazione non hanno proprietà statistiche note.

Tale approccio presenta un limite anche dal punto di vista computazionale. Nonostante il problema di minimo fornito dall'equazione enunciata possa sembrare semplice, la sua minimizzazione con metodi numerici può subire problemi di minimi locali multipli. Tale aspetto è abbastanza evidente nel caso non ponderato, ove tutti i pesi sono uguali, come nel caso del modello standard Lee-Carter. In questo caso, si sceglie il valore di β che massimizza la forma quadratica $\beta' C \beta$ nella sfera unitaria $\|\beta\| = 1$. Sfortunatamente, una volta che la forma quadratica $\beta' C \beta$ (nelle variabili A) è limitata alla sfera

unitaria, acquisisce esattamente 2A limiti locali, e quindi qualsiasi algoritmo numerico di massimizzazione che si basa solo sulle condizioni del primo ordine, può convergere in una soluzione non ottimale. E' bene ricordare tuttavia, che la scomposizione in componenti principali (SVD) per stimare β permette di oltrepassare questo problema. Quando i pesi non sono tutti uguali, SVD non può più essere utilizzato. In questo caso ciò che accade al numero di estremi locali non è ovvio, e non è più possibile dimostrare facilmente che è ancora uguale a 2A. Tuttavia, poiché la funzione da massimizzare nell'equazione dipende dai pesi d_{xt} in modo lineare, si tratta di una deformazione regolare della funzione obiettivo i cui pesi sono costanti. Di conseguenza, è ancora possibile aspettarsi un certo numero di minimi e massimi locali.

Esistono pertanto tecniche di minimizzazione che conducono solo all'apparenza a risultati significativi. Ciò aggiunge un altro motivo a favore del metodo MLE rispetto al WLS. Il primo si basa sull'osservazione che il numero di decessi sia una variabile casuale di conteggio, che può essere modellata da un processo di Poisson. In luogo delle ipotesi OLS per la log-mortalità con errori omoschedastici, utilizzati da Lee e Carter, è possibile utilizzare le seguenti specificazioni Poisson:

$$\frac{d_{xt}}{p_{xt}} \sim \text{Poisson}(e^{\mu_{xt}})$$

$$\mu_{xt} = \alpha_{xt} + \beta_x \kappa_t$$

Questo approccio porta ad uno stimatore di massima verosimiglianza standard, con le solite note proprietà: la sua unica difficoltà consiste nel fatto che molti pacchetti statistici comunemente usati per una regressione Poisson, non risulteranno in grado di gestire la forma bilineare $\beta_x \kappa_t$. Tuttavia, in un recente documento Brouhns et al. (2002) riportano l'implementazione della strategia di Wilmoth utilizzando LEM, un pacchetto software disponibile gratuitamente per l'analisi di dati categorici (Brouhns, Denuit and Vermunt, 2002). Pertanto, se si sceglie tra gli stimatori WLS e MLE, risultano preferibili i secondi, anche se è stata evidenziata solo una sottile differenza tra i due nell'analisi dei dati statunitensi.

2.2 Il modello Cairns-Blake-Dowd (M5)

Analisi empiriche sui dati di mortalità suggeriscono che il logaritmo naturale degli odds, $\ln \frac{q_{xt}}{p_{xt}}$, assume una forma lineare rispetto all'età x per un periodo temporale di t anni.

Cairns et. Al. (2006) hanno quindi proposto il seguente modello che include 2 fattori temporali:

$$\ln \frac{q_{xt}}{p_{xt}} = \kappa_t^{(1)} + (x - \bar{x}) \kappa_t^{(2)},$$

ovvero:

$$q_{xt} = \frac{\exp\left(\kappa_t^{(1)} + (x - \bar{x}) \kappa_t^{(2)}\right)}{1 + \exp\left(\kappa_t^{(1)} + (x - \bar{x}) \kappa_t^{(2)}\right)}.$$

$\kappa_t^{(1)}$ e $\kappa_t^{(2)}$ sono due processi stocastici che costituiscono una serie temporale bivariata e governano la proiezione dei tassi di mortalità.

Questo modello non pone problemi di identificazione dei parametri, ciò vuol dire che non richiede l'imposizione di vincoli.

Generalmente $\kappa_t^{(1)}$ decresce nel tempo, così come nel modello Lee-Carter, mostrando come i tassi di mortalità diminuiscano nel tempo per tutte le età.

Se durante il periodo di osservazione dei dati, si registrano incrementi di mortalità più elevati alle età giovanili rispetto che alle età anziane, allora $\kappa_t^{(2)}$ aumenta nel tempo.

Rispetto al modello di Lee-Carter, il modello Cairns-Blake-Dowd (M5), mostra cambiamenti dei tassi di mortalità non perfettamente correlati con le età.

2.2.1 Stima dei parametri

La stima dei parametri del modello viene effettuata in base al metodo dei minimi quadrati.

- Si consideri il seguente modello di regressione:

$$\ln \frac{\hat{q}_{xt}}{\hat{p}_{xt}} = \kappa_t^{(1)} + (x - \bar{x}) \kappa_t^{(2)} + \epsilon_{xt}$$

dove la variabile ϵ_{xt} rappresenta il termine di errore: errori indipendenti ed identicamente distribuiti con distribuzione $\mathcal{N}(0, \sigma_\epsilon^2)$.

- Si definisce una funzione obiettivo da minimizzare (metodo dei minimi quadrati), che fornisce la stima dei parametri temporali:

$$\vartheta_t(\kappa) = \sum_{x=x_1}^{x_m} \left(\ln \frac{\hat{q}_{xt}}{\hat{p}_{xt}} - \kappa_t^{(1)} + (x - \bar{x}) \kappa_t^{(2)} \right)^2.$$

Anche questo modello può essere calibrato utilizzando altri metodi di stima dei parametri.

2.2.2 Proiezione di $\kappa_t^{(1)}$ e $\kappa_t^{(2)}$

Per produrre previsioni sulla mortalità, si consideri per i parametri $\kappa_t^{(1)}$ e $\kappa_t^{(2)}$ la seguente dinamica

$$\begin{cases} \kappa_t^{(1)} = \kappa_{t-1}^{(1)} + d_1 + \xi_t^{(1)} \\ \kappa_t^{(2)} = \kappa_{t-1}^{(2)} + d_2 + \xi_t^{(2)} \end{cases}$$

e la seguente matrice di varianze e covarianze:

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$$

con

$$\hat{\sigma}_i^2 = \frac{1}{t_n - t_1} \sum_{t=t_2}^{t_n} \left(\hat{\kappa}_t^{(i)} - \hat{\kappa}_{t-1}^{(i)} - d_i \right)^2 \quad i = 1, 2$$

e

$$\hat{\sigma}_{12} = \frac{1}{t_n - t_1} \sum_{s=t_1}^{t_n-1} \sum_{t=t_1}^{t_n-1} \left(\hat{\kappa}_{s+1}^{(1)} - \hat{\kappa}_s^{(1)} - \hat{d}_1 \right) \left(\hat{\kappa}_{t+1}^{(2)} - \hat{\kappa}_t^{(2)} - \hat{d}_2 \right).$$

Si effettua infine la stima del drift del processo ARIMA:

$$d_i = \frac{\hat{\kappa}_{t_n}^{(i)} - \hat{\kappa}_{t_1}^{(i)}}{t_n - t_1}, \quad i = 1, 2.$$

2.3 Il modello di Renshaw-Haberman (2006): Lee-Carter con effetto coorte

In alcuni paesi si osservano tassi di mortalità che sembrano influenzati non solo da età ed anno di calendario, ma anche dall'anno di nascita della coorte.

Renshaw e Haberman (2006) generalizzano il modello Lee-Carter incorporando dunque tale effetto, in modo tale da ottenere il seguente predittore η_{xt} , e cioè $\ln m_{xt}$:

$$\eta_{xt} = \alpha_x + \beta_x^{(1)} \kappa_t^{(1)} + \beta_x^{(0)} \gamma_{t-x}$$

ove:

- $\beta_x^{(1)}$ descrive per ogni età come la mortalità reagisce al variare del parametro temporale
- $\beta_x^{(0)}$ descrive per ogni età come la mortalità reagisce al variare dell'effetto coorte
- γ_{t-x} è il parametro che rappresenta l'effetto coorte ($t - x$ rappresenta l'anno di nascita)

In questo modello le proiezioni della mortalità derivano utilizzando la previsione delle serie temporali dei $\kappa_t^{(1)}$ e γ_{t-x} stimati, considerando processi ARIMA con l'ipotesi di indipendenza tra il periodo e l'effetto coorte.

Per stimare il modello, Renshaw e Haberman (2006) ipotizzano decessi (componente casuale) distribuiti come una Poisson e l'utilizzo di una funzione log link per ottenere la forza di mortalità μ_{xt} . Come per il modello Lee-Carter, il predittore η_{xt} risulta essere invariante alla trasformazione:

$$(3) \quad \left(\alpha_x, \beta_x^{(1)}, \kappa_t^{(1)}, \beta_x^{(0)}, \gamma_{t-x} \right) \rightarrow$$

$$\left(\alpha_x + c_1 \beta_x^{(1)} + c_2 \beta_x^{(0)}, \frac{1}{c_3} \beta_x^{(1)}, c_3 \left(\kappa_t^{(1)} - c_1 \right), \frac{1}{c_4} \beta_x^{(0)}, c_4 \left(\gamma_{t-x} - c_2 \right) \right),$$

dove $c_1, c_2, c_3 \neq 0$ e $c_4 \neq 0$ sono costanti reali. L'identificabilità del modello può essere garantita utilizzando la seguente serie di vincoli di parametri:

$$\sum_x \beta_x^{(1)} = 1, \quad \sum_t \kappa_t^{(1)} = 0, \quad \sum_x \beta_x^{(0)} = 1, \quad \sum_{c=t_1-x_k}^{t_n-x_1} \gamma_c = 0,$$

cioè impostando

$$c_1 = \frac{1}{n} \sum_t \kappa_t^{(1)}, \quad c_2 = \frac{1}{n+k-1} \sum_{c=t_1-x_k}^{t_n-x_1} \gamma_c,$$

$$c_3 = \sum_x \beta_x^{(1)}, \quad c_4 = \sum_x \beta_x^{(0)},$$

nella trasformazione (3).

Renshaw e Haberman (2006) prendono in considerazione anche diverse sottostrutture del predittore; di particolare interesse è quella ottenuta ponendo $\beta_x^{(0)} = 1$,

$$\eta_{xt} = \alpha_x + \beta_x^{(1)} \kappa_t^{(1)} + \gamma_{t-x},$$

la quale è stata suggerita come caso più semplice in grado di risolvere problemi di stabilità del modello originale.

2.4 I modelli di Cairns-Blake-Dowd (M6-M7-M8) (2007)

Cairns et al. (2007) hanno proposto una variante del modello CBD originale (M5), che include sia l'effetto coorte che un altro parametro temporale, si ottiene così il seguente predittore:

$$\ln \frac{q_{xt}}{p_{xt}} = \kappa_t^{(1)} + (x - \bar{x}) \kappa_t^{(2)} + ((x - \bar{x})^2 - \hat{\sigma}_x^2) \kappa_t^{(3)} + \gamma_{t-x},$$

dove $\hat{\sigma}_x^2$ è il valore medio di $(x - \bar{x})^2$. Questo modello, solitamente chiamato M8, è invariante rispetto la trasformazione

$$\left(\kappa_t^{(1)}, \kappa_t^{(2)}, \kappa_t^{(3)}, \gamma_{t-x} \right) \rightarrow \left(\kappa_t^{(1)} + \phi_1 + \phi_2 (t - \bar{x}) + \phi_3 ((t - \bar{x})^2 + \hat{\sigma}_x^2), \right.$$

$$\left. \kappa_t^{(2)} - \phi_2 - 2\phi_3 (t - \bar{x}), \kappa_t^{(3)} + \phi_3, \gamma_{t-x} - \phi_1 - \phi_2 (t - x) - \phi_3 (t - x)^2 \right)$$

per le costanti reali ϕ_1, ϕ_2 e ϕ_3 , il predittore rimane invariato. Questo modello quindi, pone problemi di identificazione dei parametri. Per ovviare a tale problema i parametri sono trasformati utilizzando fattori di trasformazione che li rendono individuabili.

Il modello Cairns et al. (2009) impone l'insieme di vincoli:

$$\sum_{c=t_1-x_k}^{t_n-x_1} \gamma_c = 0, \quad \sum_{c=t_1-x_k}^{t_n-x_1} c\gamma_c = 0, \quad \sum_{c=t_1-x_k}^{t_n-x_1} c\gamma_c^2 = 0,$$

i quali assicurano fluttuazioni dell'effetto coorte attorno allo zero e che questo non abbia alcun trend quadratico o lineare percepibile. Seguendo l'approccio di Haberman e Renshaw, tali vincoli possono essere imposti attraverso l'applicazione della trasformazione con le costanti ϕ_1, ϕ_2 e ϕ_3 ottenute con la regressione γ_{t-x} su $t-x$ e $(t-x)^2$, nel seguente modo

$$\begin{aligned} \gamma_{t-x} &= \phi_1 + \phi_2(t-x) + \phi_3(t-x)^2 + \epsilon_{t-x}, \\ \epsilon_{t-x} &\sim \mathcal{N}(0, \sigma^2) \quad i.i.d \end{aligned}$$

Cairns et al. (2009) considerano inoltre altre due strutture più semplici:

$$\begin{aligned} \ln \frac{q_{xt}}{p_{xt}} &= \kappa_t^{(1)} + (x - \bar{x}) \kappa_t^{(2)} + \gamma_{t-x}, \\ \ln \frac{q_{xt}}{p_{xt}} &= \kappa_t^{(1)} + (x - \bar{x}) \kappa_t^{(2)} + (x_c - x) \gamma_{t-x}, \end{aligned}$$

dove x_c è un parametro costante da stimare. Queste strutture sono tipicamente riconosciute come modelli M6 e M7, rispettivamente.

2.5 Il modello Plat (2009)

Plat (2009) combina il modello CBD con alcune caratteristiche del modello Lee-Carter per ottenere un modello che si adatta bene a tutte le età e cattura l'effetto coorte. La struttura del predittore proposta, assume l'esistenza di una funzione statica dell'età α_x , tre termini age-period con parametri predefiniti $\beta_x^{(1)} = 1, \beta_x^{(2)} = \bar{x} - x, \beta_x^{(3)} = (\bar{x} - x)^+ = \max(0, \bar{x} - x)$, e di un effetto coorte con parametri predefiniti $\beta_x^{(0)} = 1$.

Pertanto, il predittore è dato da:

$$\ln m_{xt} = \alpha_x + \kappa_t^{(1)} + (x - \bar{x}) \kappa_t^{(2)} + (x - \bar{x})^+ \kappa_t^{(3)} + \gamma_{t-x}.$$

ove i parametri del modello vengono stimati ipotizzando una distribuzione dei decessi di tipo Poisson. Le seguenti trasformazioni dei parametri, lasciano il predittore invariato:

$$\left(\alpha_x, \kappa_t^{(1)}, \kappa_t^{(2)}, \kappa_t^{(3)}, \gamma_{t-x} \right) \rightarrow \left(\alpha_x + \phi_1 - \phi_2 x + \phi_3 x^2, \kappa_t^{(1)} + \phi_2 t + \phi_3 (t^2 - 2\bar{x}t), \right.$$

$$\left. \kappa_t^{(2)} + 2\phi_3 t, \kappa_t^{(3)}, \gamma_{t-x} - \phi_1 - \phi_2(t - x) - \phi_3 (t - x)^2 \right)$$

$$\left(\alpha_x, \kappa_t^{(1)}, \kappa_t^{(2)}, \kappa_t^{(3)}, \gamma_{t-x} \right) \rightarrow \left(\alpha_x + c_1 + c_2 (\bar{x} - x) + c_3 (\bar{x} - x)^+, \right.$$

$$\left. \kappa_t^{(1)} - c_1, \kappa_t^{(2)} - c_2, \kappa_t^{(3)} - c_3, \gamma_{t-x}, \right)$$

dove $c_1, c_2, c_3, \phi_1, \phi_2$, e ϕ_3 sono delle costanti reali. Al fine di garantire l'identificabilità dei parametri è necessario, anche in questo caso, imporre il seguente insieme di vincoli:

$$\sum_t \kappa_t^{(1)} = 0, \sum_t \kappa_t^{(2)} = 0, \sum_t \kappa_t^{(3)} = 0,$$

$$\sum_{c=t_1-x_k}^{t_n-x_1} \gamma_c = 0, \sum_{c=t_1-x_k}^{t_n-x_1} c\gamma_c = 0, \sum_{c=t_1-x_k}^{t_n-x_1} c^2\gamma_c = 0$$

I primi tre vincoli assicurano che gli indici del periodo siano centrati attorno allo zero, mentre gli ultimi tre assicurano che l'effetto coorte fluttui attorno allo zero e che non abbia una tendenza lineare o quadratica. Seguendo

il modello di Haberman e Renshaw, i vincoli sull'effetto coorte possono essere imposti applicando la prima trasformazione con le costanti ϕ_1, ϕ_2 e ϕ_3 ottenuta attraverso la regressione di γ_{t-x} su $(t-x)^2$, in modo che

$$\gamma_{t-x} = \phi_1 + \phi_2(t-x) + \phi_3(t-x)^2 + \epsilon_{t-x},$$

$$\epsilon_{t-x} \sim \mathcal{N}(0, \sigma^2) \quad i.i.d$$

I vincoli sugli indici di periodo invece, possono essere imposti applicando la seconda trasformazione con

$$c_i = \frac{1}{n} \sum \kappa_t^{(i)}, \quad i = 1, 2, 3.$$

Nei casi in cui interessano solo le età più anziane, Plat (2009) suggerisce di abbandonare dal predittore il terzo termine di periodo:

$$\ln m_{xt} = \alpha_x + \kappa_t^{(1)} + (\bar{x} - x) \kappa_t^{(2)} + \gamma_{t-x}.$$

Si noti che questa versione del modello Plat ridotto, pone gli stessi problemi di identificabilità del modello completo, considerando però l'omissione delle trasformazioni e dei vincoli concertanti $\kappa_t^{(3)}$ e c_3 .

Capitolo 3

Un modello di deep learning: le reti neurali

Nell'ambito dei modelli estrapolativi di mortalità, il modello Lee-Carter (1992) rappresenta quello più frequentemente utilizzato. Esso ha il vantaggio, rispetto ad altri, di esser caratterizzato da un piccolo numero di parametri e da un buon grado di robustezza. La parsimonia del modello si traduce in un metodo in cui la mortalità risulta stabile con l'età.

Negli ultimi anni, grazie al miglioramento delle capacità computazionali, sono tornate oggetto di grande interesse le tecniche di machine learning. Il machine learning racchiude molti algoritmi i quali apprendono a partire da grandi insiemi di dati, caratterizzati anche da numerose variabili. Si tratta di tecniche molto utili in quanto permettono di rilevare modelli altrimenti sconosciuti e non identificabili. Nonostante l'utilizzo sempre crescente della ricerca nei diversi ambiti, le applicazioni di machine learning non sono molto popolari nel campo demografico, infatti i risultati ottenuti sono spesso considerati delle "black boxes" difficili da interpretare. Inoltre, gli algoritmi di machine learning si basano su dati senza il supporto di ipotesi specifiche, al quale invece i demografi prestano particolare interesse.

I contributi per la ricerca in quest'ambito sono, tuttavia, in forte aumento.

Le reti neurali artificiali fanno parte di una classe più ampia di modelli formulati per lo studio di sistemi complessi, a dinamica non lineare, di tipo caotico, e così via. Questi modelli sono stati introdotti nei settori di ricerca più innovativi in diversi ambiti disciplinari. Il loro carattere generale ed astratto li rende applicabili a fenomeni molto diversi tra loro, e quindi di inte-

resse potenziale molto ampio, e anche a fenomeni sociali, dal comportamento nei mercati finanziari alle dinamiche demografiche.

L'elaborato si basa su un nuovo approccio proposto da Hainaut (2018), in cui viene utilizzata una rete neurale al fine di prevedere e simulare i tassi di mortalità futuri. L'autore utilizza un analizzatore neurale per identificare i processi temporali latenti e per predire direttamente la mortalità. Questo metodo consente di rilevare e duplicare le non linearità osservate nell'evoluzione delle forze di mortalità logaritmiche. Deprez et al. (2017) utilizzano alcune tecniche di machine learning per migliorare la stima dei tassi di log-mortalità, le quali vengono successivamente estese da Levantesi e Pizzorusso (2018) per prevedere la mortalità considerando la struttura del modello Lee-Carter. L'elaborato farà riferimento dunque, ad una tecnica di deep learning finalizzata al miglioramento della capacità predittiva del modello Lee-Carter, tradizionalmente utilizzato nella letteratura.

3.1 Modello della rete neurale

Il termine rete neurale (NN) è nato come un modello matematico in grado di replicare le reti neurali del cervello umano (Minsky e Pitts (1943), Wiener (1948)).

Le reti neurali artificiali sono algoritmi di analisi dati e di modellazione numerica. Vengono dette “data driven models”, ossia modelli di calcolo basati sul processamento di dati al fine di testare la sensibilità di un certo fenomeno ad alcuni parametri.

Le NN sono particolarmente utili nelle analisi di fenomeni in cui il numero di parametri in considerazione è molto elevato, e le relazioni tra essi sono fortemente non lineari e molto complesse. Dal punto di vista della tipologia di modellazione, le NN costituiscono una “black-box”, dunque l'algoritmo di calcolo prescinde completamente dallo studio del fenomeno fisico che sta alla base della relazione di causa-effetto tra dati di input e di output; pertanto, le reti neurali sono in grado di elaborare i dati sperimentali a disposizione (input) e produrre delle predizioni dei valori desiderati (output), ricostruendo artificialmente gli effetti della complessa relazione di causa-effetto incognita.

Essenziale, ai fini di una buona modellazione mediante reti neurali, è il poter disporre di un database numericamente consistente e affidabile di valori sperimentali. L'accuratezza e attendibilità dei risultati di una rete neurale dipendono infatti proprio dalla precisione dei dati di input, che non fungono

solo da punto di partenza del modello, ma costituiscono lo strumento interno di calibrazione e validazione del modello stesso.

Le reti neurali artificiali sono un algoritmo utilizzato per risolvere problemi di natura complessa non facilmente codificabili, e sono una colonna portante del machine learning come viene inteso oggi. L'architettura della rete neurale include neuroni, connessioni sinaptiche che collegano i neuroni e algoritmi di apprendimento. Sono chiamate "reti neurali" perché il comportamento dei nodi che le compongono ricorda vagamente quello dei neuroni biologici. Un neurone riceve in ingresso segnali da vari altri neuroni tramite connessioni sinaptiche e li integra. Se l'attivazione che ne risulta supera una certa soglia genera un potenziale d'azione che si propaga attraverso il suo assone a uno o più neuroni.

Le reti neurali artificiali sono costituite da molte unità di elaborazione (dette unità o nodi o neuroni artificiali), di solito ordinate in strati, dal funzionamento molto semplice e tra loro interconnesse. Nella rete così costituita si fa passare un segnale (in forma di esempi, anche detti patterns) che attiva o inibisce le unità. Esse, con opportune regole matematiche, trasferiscono il segnale ad altre unità fino a produrre un output quantitativo. In altre parole, una unità riceve attivazione (o inibizione) dalle unità da cui arrivano connessioni e, a sua volta, manda attivazione (o inibizione) alle unità verso le quali ha delle connessioni. L'elaborato si occuperà prevalentemente di un tipo di reti, dette feedforward, che hanno le unità disposte su almeno tre strati e connessioni di tipo unidirezionale tra ogni unità di uno strato e tutte le altre unità dello strato successivo. L'attivazione o inibizione che arriva ad una certa unità attraverso le altre unità cui è connessa, dipende dai pesi che caratterizzano i legami. Un peso su un legame alto fa passare molta attivazione; un peso basso fa passare poca attivazione. Un peso positivo trasmette attivazione, mentre un peso negativo inibizione. Ciò che caratterizza le reti neurali artificiali è il loro funzionamento in parallelo: ogni nodo della rete costituisce una unità di elaborazione autonoma che effettua dei calcoli matematici in parallelo con tutte le altre.

Una rete neurale è composta da diversi «layer» (strati) di nodi, ciascuno dei quali è collegato ai nodi del layer successivo. Tipicamente le reti neurali sono composte da almeno tre tipi di strati, ognuno dei quali contiene diversi neuroni:

- Un *input layer*, ovvero i dati di ingresso

- Uno o più *hidden layer*, dove avviene l'elaborazione vera e propria. Sono opzionali e partecipano alle comunicazioni all'interno della rete
- Un *output layer*, contenente il risultato finale.

Nella rete ciascuna unità ottiene delle informazioni "ponderate" attraverso i collegamenti sinaptici dalle altre connesse, e restituisce un output utilizzando una funzione di attivazione che trasforma la somma ponderata dei segnali in input. Come per i neuroni biologici, i nodi delle NN integrano i segnali in ingresso secondo una apposita funzione. Se il valore di questa operazione supera la soglia prevista, il nodo sollecitato trasmette questo valore al layer successivo, altrimenti il valore trasmesso sarà zero.

Considerando un singolo neurone H — chiamato perceptron da Rosenblatt (1958) — l'output corrispondente è definito come segue:

$$\text{out}_H = \phi(\mathbf{w}^T \mathbf{x} + b),$$

dove $\mathbf{x} \in \mathbb{R}^d$ è l'input, $\mathbf{w} \in \mathbb{R}^d$ è il peso sinaptico associato, $d \in \mathbb{N}$ è il numero di segnali in input, e ϕ è la funzione di attivazione. Il termine $b \in \mathbb{R}$ rappresenta l'errore, anche chiamato ingresso di attivazione. È importante notare che la funzione ϕ deve essere differenziabile perché le equazioni di apprendimento sono basate sul gradiente.

3.2 Caratteristiche generali

Le reti neurali artificiali presentano alcune caratteristiche che si rivelano interessanti in molti campi di ricerca e domini di applicazione. Benché molte di queste caratteristiche varino da modello a modello, ve ne sono alcune sufficientemente generali.

Robustezza Una rete neurale è resistente al rumore, ovvero è in grado di continuare a dare una risposta corretta anche se alcune delle sue connessioni vengono eliminate o lesionate o se viene aggiunto del rumore al segnale d'ingresso, ai canali di trasmissione o alla funzione di attivazione dei nodi. Questa proprietà è comune anche ai sistemi nervosi biologici dove la capacità di apprendere e ricordare non viene alterata in modo sostanziale dalla perdita continua di neuroni. In caso, le reti lesionate possono essere addestrate nuovamente per riacquistare le abilità perse.

Flessibilità Un modello neurale può essere impiegato per un grande numero di finalità diverse: esso non ha bisogno di conoscere le proprietà del dominio specifico di applicazione perché le apprende in base all'esperienza. Questo non significa che un qualsiasi modello neurale possa essere utilizzato per tutti i tipi di compiti, ma implica che l'utente non deve necessariamente conoscere le soluzioni dettagliate e analitiche che caratterizzano il problema sotto indagine. In generale l'utente di una rete neurale deve essere in grado di individuare precisamente le finalità del progetto, il tipo di compito e una serie di vincoli al fine di valutare qual è il modello neurale che risulta più appropriato.

Generalizzazione Una rete neurale che è stata addestrata su un numero limitato di esempi è in grado di produrre una risposta adeguata a dei pattern d'ingresso che non ha mai visto in precedenza, ma che presentano tuttavia qualche somiglianza con gli esempi presentati durante la fase di addestramento. Questa proprietà deriva in parte dal fatto che molti modelli neurali rappresentano internamente un numero di associazioni stimolo-risposta più grande del numero di sinapsi disponibili; nel far questo la rete neurale tende ad estrarre le caratteristiche invarianti dei pattern d'ingresso piuttosto che memorizzare ciascun singolo pattern. La capacità di generalizzare a nuovi stimoli è una caratteristica molto apprezzata nei tipici campi di applicazione delle reti neurali dove spesso è impossibile ottenere una collezione esaustiva di tutti i dati su cui la rete neurale dovrà operare. Per questi motivi, le ANN vengono impiegate in molti settori di ricerca ed in molti campi di applicazione, dall'informatica alla meccanica, dalla biomedica alla finanza, dalla matematica alla chimica. E' possibile utilizzare una rete neurale artificiale per l'approssimazione di funzioni non lineari, regressione e predizione, per la classificazione in categorie (clustering), per il data mining, nei sistemi di controllo nel campo della robotica e nel controllo numerico, nel riconoscimento di sequenze e di modelli (immagini, caratteri o oggetti in generale).

3.2.1 Tipi di architetture della rete

Esistono molti tipi di reti neurali che sono differenziati sulla base di alcune caratteristiche fondamentali:

- tipo di utilizzo
- tipo di apprendimento

- algoritmo di apprendimento
- architettura dei collegamenti.

La tipologia di architettura di rete è strettamente dipendente dall'algoritmo di apprendimento che si vuole utilizzare. In relazione alla maniera in cui le unità sono collegate dalle sinapsi, vengono definite diverse tipologie di reti. L'importante svolta si ebbe nel 1958 con Frank Rosenblatt, il quale presentò il primo schema di rete neurale, il Percettrone: tale modello costituisce la forma più semplice di rete neurale usata per classificare pattern linearmente separabili, ovvero piani collocati ai lati opposti di un iperpiano. Il Percettrone consiste in un singolo neurone con soglia e pesi modificabili. L'algoritmo di apprendimento si basa sul Teorema di convergenza del percettrone ed il modello di rete è feedforward (gli impulsi si propagano in un'unica direzione, come sarà illustrato successivamente). La restrizione del Percettrone era il campo di applicazione: esso infatti si limitava a riconoscere forme, classificarle in gruppi separati e calcolare semplici funzioni (come gli operatori booleani AND, OR e NOT, poichè linearmente separabili). Il passo successivo a questo modello fu il Percettrone Multistrato (MLP), in grado di calcolare funzioni non linearmente separabili (come lo XOR).

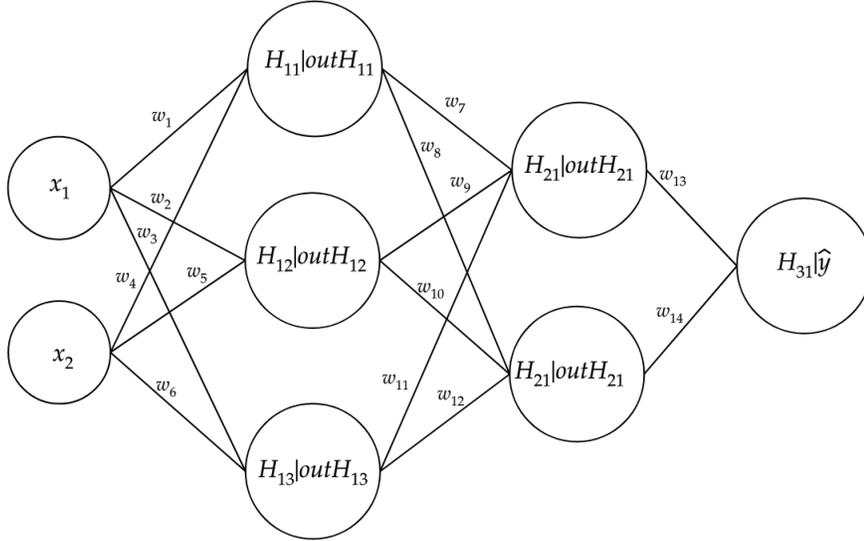
Nel modello classico della rete neurale (feedforward NN), che verrà considerato in seguito, l'informazione si sposta dall'input all'output layer in un'unica direzione; ciò vuol dire che il segnale di input si propaga attraverso la rete in avanti da layer a layer. Ciò non accade nelle reti neurali ricorrenti (RNN), in cui le informazioni vengono elaborate ciclicamente utilizzando delle sinapsi aggiuntive in modo tale da riproporre come input il risultato del processo di elaborazione.

In generale si identificano tre classi di reti:

1. *Reti feedforward ad uno strato*: la più semplice rete feedforward è il percettrone a singolo strato, dall'inglese SingleLayer Perceptron (SLP), utilizzato verso la fine degli anni '60. Un SLP è costituito da un strato in ingresso, seguito direttamente dall'uscita. Ogni unità di ingresso è collegata ad ogni unità di uscita. Questo tipo di rete neurale ha un solo strato che effettua l'elaborazione dei dati, e non presenta nodi nascosti, da cui il nome. Gli SLP sono tuttavia molto limitati a causa del piccolo numero di connessioni e dell'assenza di gerarchia nelle caratteristiche che la rete può estrarre dai dati, ragione per cui è in grado di combinare i dati in ingresso una sola volta.

2. *Reti feedforward a più strati*: grazie all'introduzione di un MultiLayer Perceptron (MLP) sono state ispirate per le reti neurali nuove strutture di maggiore potenza. Questa classe di reti feedforward si distingue dalla precedente dal fatto che tra lo strato di input e quello di output vi sono uno o più strati di neuroni nascosti (hidden layers). Ogni strato ha connessioni entranti dal precedente strato e uscenti in quello successivo, quindi la propagazione del segnale avviene in avanti senza cicli e senza connessioni trasversali. Questo tipo di architettura fornisce alla rete una prospettiva globale in quanto aumentano le interazioni tra neuroni. Nella MultiLayer Perceptron, i neuroni sono dunque predisposti su diversi strati e ciascuna unità è completamente collegata a quelle del livello precedente.
3. *Reti ricorrenti o feedback*: una rete ricorrente si distingue dalle precedenti nel fatto che è ciclica. La presenza di cicli ha un impatto profondo sulle capacità di apprendimento della rete e sulle sue performance, in particolare rendono il sistema dinamico.

La figura sottostante mostra la rappresentazione tipica di una rete neurale feedforward a più strati. Si tratta di un sistema dinamico avente la topologia di un grafo orientato con nodi, i neuroni artificiali, ed archi, i quali rappresentano i pesi sinaptici. I cerchi rappresentano i neuroni e le linee rappresentano le connessioni sinaptiche. Nell'algoritmo queste connessioni vengono "pesate" tramite fattori moltiplicativi, che descrivono la "forza" della connessione stessa. Come per i neuroni biologici, i nodi della NN integrano i segnali in ingresso secondo una apposita funzione.



L'output $\mathbf{outH} \in \mathbb{R}^{n_h}$ di un generico hidden layer con n_h neuroni, è definito come segue:

$$\mathbf{outH} = \phi (W^T \mathbf{x} + \mathbf{b})$$

dove $W \in \mathbb{R}^{d \times n_h}$ rappresenta la matrice dei pesi e $\mathbf{b} \in \mathbb{R}^{n_h}$ è il vettore degli errori. Nello schema di un MultiLayer Perceptron, l'output di un hidden layer diviene l'input per il layer successivo.

Si consideri un problema di regressione, ove $g \in \mathbb{N}$ è il numero degli hidden layers, l'output $\hat{y} \in \mathbb{R}$ si ottiene da:

$$\mathbf{outH}_1 = \phi_1 (W_1^T \mathbf{x} + \mathbf{b}_1),$$

$$\mathbf{outH}_2 = \phi_2 (W_2^T \mathbf{outH}_1 + \mathbf{b}_2),$$

...

$$\hat{y} = \phi_g (W_g^T \mathbf{outH}_g + \mathbf{b}_g),$$

dove W_1, W_2, \dots, W_g sono le matrici peso, $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_g$ sono i vettori degli errori, e $\phi_1, \phi_2, \dots, \phi_g$ sono le funzioni di attivazione non necessariamente diverse l'una dall'altra. E' interessante notare che la dimensione delle matrici peso e dei vettori degli errori, dipende dal numero di unità presenti negli

CAPITOLO 3. UN MODELLO DI DEEP LEARNING: LE RETI NEURALI 40

strati nascosti. Aumentando il numero di hidden layers, aumenta il livello di astrazione dei dati in input.

Dal punto di vista del tipo di utilizzo possiamo distinguere tre categorie basilari:

- Memorie associative: possono apprendere associazioni tra patterns (insieme complesso di dati come l'insieme dei pixels di una immagine) in modo che la presentazione di un pattern A dia come output il pattern B anche se il pattern A è impreciso o parziale (resistenza al rumore). Esiste anche la possibilità di utilizzare la memoria associativa per fornire in uscita il pattern completo in risposta ad un pattern parziale in input.
- Simulatori di funzioni matematiche: sono in grado di comprendere la funzione che lega output con input in base a degli esempi forniti in fase di apprendimento. Dopo la fase di apprendimento, la rete è in grado di fornire un output in risposta ad un input anche diverso da quelli usati negli esempi di addestramento. Ne consegue una capacità della rete di interpolazione ed estrapolazione sui dati del training set. Tale capacità è facilmente verificabile addestrando una rete con una sequenza di dati input/output proveniente da una funzione nota e risulta, invece, utile proprio per il trattamento e la previsione di fenomeni di cui non sia chiaro matematicamente il legame tra input e output. In ogni caso la rete si comporta come una “black box”, poichè non svela in termini leggibili la funzione di trasferimento che è contenuta al suo interno. Di questo tipo fa parte la rete a retropropagazione dell'errore o error back propagation che è quella attualmente più utilizzata per efficacia e flessibilità.
- Classificatori: con essi è possibile classificare dei dati in specifiche categorie in base a caratteristiche di similitudine. In questo ultimo tipo di rete esiste il concetto di apprendimento non supervisionato o “autoorganizzante”, nel quale i dati di input vengono distribuiti su categorie non predefinite. L'algoritmo di apprendimento di una rete neurale dipende essenzialmente dal tipo di utilizzo della stessa, così come l'architettura dei collegamenti. Le reti multistrato prevedono ad esempio l'algoritmo a retropropagazione dell'errore o sono addestrate tramite algoritmi genetici.

Gli algoritmi di apprendimento utilizzati per istruire le reti neurali sono infine divisi in tre categorie. La scelta di quale usare dipende dal campo di applicazione per cui la rete è progettata e dalla sua tipologia (feedforward o feedback). Gli algoritmi sono:

- Apprendimento supervisionato
- Apprendimento non supervisionato
- Apprendimento di rinforzo

Apprendimento supervisionato Nell'apprendimento supervisionato si fornisce alla rete un insieme di input ai quali corrispondono output noti (training set). Analizzandoli, la rete apprende la relazione che li unisce. In tal modo impara a generalizzare, ossia a calcolare nuove associazioni corrette input-output processando input esterni al training set. Man mano che la macchina elabora output, si procede a correggerla per migliorarne le risposte variando i pesi. Ovviamente, aumentano i pesi che determinano gli output corretti, diminuiscono quelli che generano valori non validi. Il meccanismo di apprendimento supervisionato impiega quindi l'Error Back-Propagation, ma è molto importante l'esperienza dell'operatore che istruisce la rete. Il motivo risiede nel complesso compito di trovare un rapporto adeguato fra le dimensioni del training set, quelle della rete e l'abilità a generalizzare che si tenta di ottenere. Un numero eccessivo di parametri in ingresso e una troppo potente capacità di elaborazione, paradossalmente, rendono difficile alla rete neurale imparare a generalizzare, perché gli input esterni al training set vengono valutati dalla rete come troppo dissimili ai sofisticati e dettagliati modelli che conosce. D'altro canto, un training set con variabili scarse porta per la via opposta alla stessa conclusione: la rete, in questo caso, non ha sufficienti parametri per apprendere a generalizzare. Il giusto compromesso, insomma, è un compito che necessita di molta preparazione ed esperienza. Le reti feedforward come il MLP utilizzano l'apprendimento supervisionato.

Apprendimento non supervisionato In una rete neurale ad apprendimento non supervisionato, la medesima riceve solo un insieme di variabili di input. Analizzandole, la rete deve creare dei cluster rappresentativi per categorizzarle. Anche in questo caso i valori dei pesi è dinamico, ma sono i nodi stessi a modificarli. Esempi di reti ad apprendimento supervisionato sono SOM e la rete di Hopfield.

Apprendimento di rinforzo Nelle reti neurali che apprendono mediante l'algoritmo di rinforzo, non esistono né associazioni input-output di esempi, né un aggiustamento esplicito degli output da ottimizzare. I circuiti neurali imparano esclusivamente dall'interazione con l'ambiente. Su di esso, eseguono una serie di azioni. Dato un risultato da ottenere, è considerato rinforzo l'azione che avvicina al risultato; viceversa, la rete apprende a eliminare le azioni negative, ossia foriere di errore. Detto in altri termini, un algoritmo di apprendimento per rinforzo mira ad indirizzare la rete neurale verso il risultato sperato con una politica di incentivi (azioni positive) e disincentivi (azioni negative).

3.3 Back-propagation

Nel 1986 Rumelhart, Hinton e Williams propongono un potente algoritmo di apprendimento che divenne presto uno tra i metodi più efficienti per l'addestramento di una rete neurale: la retropropagazione dell'errore, in inglese backward propagation of errors (solitamente abbreviato in backpropagation). Tale tecnica consiste nel confrontare iterativamente la prestazione fornita dalla rete e quella desiderata in risposta ad una determinata combinazione di stimoli di ingresso, e nel modificare di conseguenza le interconnessioni, fino a ottenere la prestazione voluta. Ciò permette di perfezionare in stadi successivi l'apprendimento automatico di una rete neurale. Con la retropropagazione dell'errore si addestra la rete ad eseguire un determinato compito in risposta a certi stimoli. Non si definiscono a priori le connessioni della rete, ma queste vengono a modificarsi progressivamente, in modo automatico, fino a che la rete non è grado di compiere il compito richiesto. La retropropagazione dell'errore rappresenta dunque un potente algoritmo per l'allenamento delle reti neurali artificiali, il quale regolando ripetutamente i pesi delle connessioni nella rete, ha lo scopo di ridurre al minimo la differenza tra il vettore di output ottenuto dal modello e il vettore di output desiderato. Esso comporta un problema di ottimizzazione senza limiti, il cui obiettivo è quello di minimizzare una funzione nell'elevato spazio dimensionale. Si definisca la seguente funzione perdita:

$$E = \frac{\sum (y - \hat{y})^2}{2}$$

la quale misura la differenza tra i valori previsti \hat{y} e quelli osservati y . La quantità E dipende dalle matrici dei pesi W_1, W_2, \dots, W_g che influenzano i

CAPITOLO 3. UN MODELLO DI DEEP LEARNING: LE RETI NEURALI 43

valori di \hat{y} . L'obiettivo del back-propagation consiste nel trovare i valori dei pesi sinaptici che minimizzano la quantità E .

Tra molti altri algoritmi, il back-propagation è il più utilizzato per l'addestramento delle reti neurali feedforward. Si tratta di un algoritmo di apprendimento che confronta i valori previsti con quelli desiderati (obiettivi). Sulla base della differenza così calcolata, l'algoritmo modifica i pesi sinaptici della rete neurale, facendo convergere progressivamente il set dei valori di uscita, ovvero degli output, verso quelli desiderati. L'algoritmo confronta i valori previsti con quelli desiderati e modifica i pesi sinaptici propagando indietro il gradiente della funzione perdita. Schematicamente, la procedura alterna passaggi di propagazione avanti e indietro:

- nella fase forward, la previsione \hat{y} viene calcolata aggiustando i pesi sinaptici,
- nella fase backward, i pesi vengono regolati in modo da ridurre l'errore E della rete.

La rete neurale esegue iterativamente la propagazione avanti e indietro e modifica i pesi così da trovare la combinazione che minimizza la funzione di perdita.

Analiticamente, l'algoritmo di back-propagation aggiorna i pesi W_g nell'ultimo layer usando la regola del delta (Bryson and Ho (1969)), definita come segue:

$$\Delta W_g = -r \frac{\partial E}{\partial W_g},$$

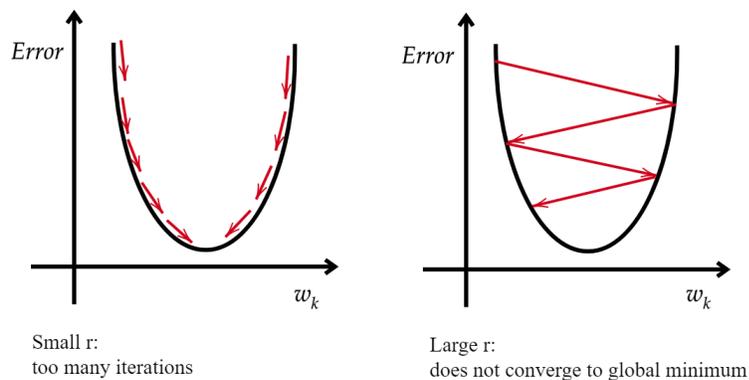
dove r è il tasso di apprendimento. Per i layers precedenti, si procede utilizzando la regola di derivazione della catena. Gli aggiornamenti della matrice dei pesi W_{g-1} vengono così calcolati:

$$\Delta W_{g-1} = -r \frac{\partial E}{\partial \text{outH}_{g-1}} * \frac{\partial \text{outH}_{g-1}}{\partial W_{g-1}}$$

e così via per gli altri layers. In modo figurato, l'idea alla base della discesa del gradiente è simile alla «discesa da una collina» fino a raggiungere un minimo globale o locale. Ad ogni aggiornamento, la ricerca si sposta nella direzione opposta del gradiente e la pendenza del gradiente insieme al tasso di apprendimento determinano l'ampiezza di questo movimento. La scelta del tasso di apprendimento influenza molto il comportamento dell'algoritmo: la

CAPITOLO 3. UN MODELLO DI DEEP LEARNING: LE RETI NEURALI 44

scelta di valori piccoli, comporta troppe iterazioni e fa sì che la convergenza sarà lenta, mentre con la scelta di un elevato valore di r si rischia di avere una rete instabile con un comportamento oscillatorio, impedendo la convergenza al minimo globale.



Le scelte relative al tipo di architettura (ad esempio il numero di hidden layers e il numero di neuroni per ciascun layer) e i valori degli iperparametri (ad esempio la velocità di apprendimento, la funzione di attivazione, il numero di epoche) rimangono un problema euristico per gli utenti delle NN: essa dipende spesso dal tipo di dati a disposizione e rappresenta una decisione non semplice.

Capitolo 4

Stima della superficie di mortalità

La forma dell'aspettativa di vita ha sempre suscitato grande interesse nei ricercatori. Dall'inizio del nuovo millennio, la sua evoluzione ha portato gli studiosi a dare maggiore enfasi alle indagini riguardo tale fenomeno.

L'aumento della longevità umana durante il ventesimo secolo, ha suscitato un'attenzione sempre maggiore verso la modellizzazione e la proiezione dei tassi di mortalità e dell'aspettativa di vita, sia da parte dei demografi che degli attuari. Sono diverse le prove empiriche che hanno ripetutamente smentito l'ipotesi di un confine imminente della vita umana. Nel loro lavoro principale, Oeppen e Vaupel coniano la cosiddetta "best practice life expectancy "(BPLE), riferita alla massima aspettativa di vita osservata tra le popolazioni nazionali durante un determinato anno civile. Le aspettative di vita alla nascita sono aumentate in modo quasi lineare ed ad una ritmo costante nel tempo, iniziando in Scandinavia intorno al 1840 e continuando da allora ad un ritmo di circa 0,24 anni all'anno per le femmine e 0,22 anni all'anno per i maschi. Il concetto di BPLE ha portato ad un maggiore interesse verso i metodi basati sull'estrapolazione dell'aspettativa di vita: oltre ad offrire un elevato grado di accuratezza previsionale, permettono l'utilizzo di una sola serie temporale univariata.

Seguendo questa linea di ricerca, l'analisi viene effettuata utilizzando l'aspettativa di vita alla nascita come input per gli anni Y_i . Pertanto, come output, si otterrà la superficie di mortalità per gli stessi anni.

La stima dei tassi futuri di mortalità attraverso l'utilizzo di misure demografiche riassuntive, risulta molto più intuitiva rispetto ai modelli tradizionali basati sull'estrapolazione dei tassi per età specifiche. Tuttavia la ricostruzio-

ne della mortalità siffatta, non risulta essere semplice. Al contrario, in molti casi, si rischia di compromettere l'accuratezza delle stime.

4.1 Dati

Le analisi sono state effettuate utilizzando i dati forniti dallo Human Mortality Database relativi ad Italia, Usa, Giappone, Australia e Russia, facendo riferimento ad entrambi i sessi. Gli anni presi in esame sono quelli che vanno dal 1950 al 2014 per ciascun paese, ad eccezione della Russia per la quale non vi sono dati disponibili prima del 1959.

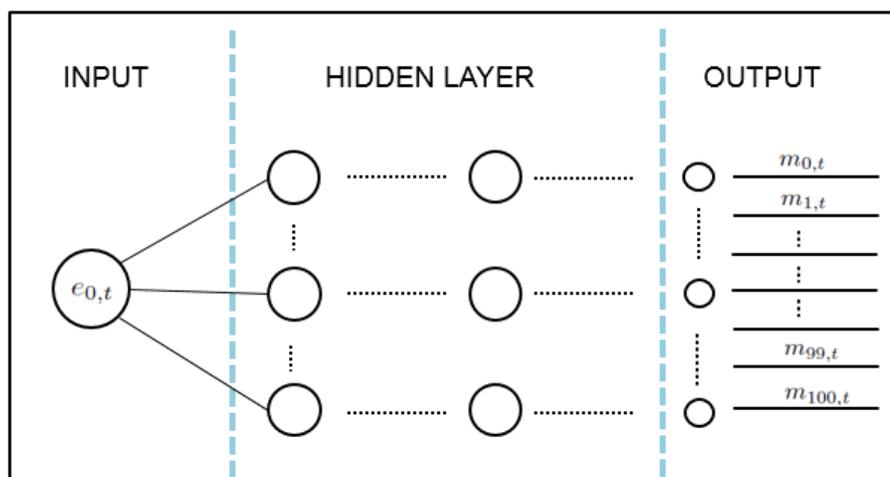
Le osservazioni sono state disposte su due matrici diverse X e Y , ove le colonne rappresentano i tassi per ogni età e le righe si riferiscono ad anni di calendario diversi. La matrice X è composta dai dati di input, nel nostro caso di studio l'aspettativa di vita alla nascita per ogni anno. La matrice Y contiene invece l'output desiderato, ovvero la matrice dei tassi di mortalità per ogni età ed anno di calendario. Per ottenere il Train - Test split, si esegue un campionamento casuale sulle righe della matrice (anni), in questo modo si evita la dipendenza temporale per anno. Considerando i dati relativi agli anni che vanno dal 1950 al 2005, l'80% del database viene utilizzato per il training ed il 20% per testare il modello.

4.2 Il modello

La rete neurale è un modello matematico prodotto dall'intelligenza artificiale, ispirato alla struttura ed al funzionamento del sistema nervoso umano. Ogni unità in una rete ottiene dagli altri connessi, tramite i collegamenti sinaptici, delle informazioni «pesate» e restituisce un output utilizzando una funzione di attivazione che trasforma la somma pesata dei segnali in ingresso. Nello specifico, per lo sviluppo delle analisi viene utilizzata la rete neurale profonda, cioè una rete di tipo feedforward: le unità sono disposte su almeno tre strati e le connessioni tra ogni neurone e tutti gli altri neuroni dello strato successivo sono di tipo unidirezionale.

Nel nostro caso la rete cattura i dati di input, apprende il modello nascosto nei dati, e restituisce l'output. A tal fine si necessita dell'aspettativa di vita alla nascita per ogni anno, come dati di input. La rete fornisce pertanto i tassi di mortalità per gli stessi anni.

La figura sottostante mostra una chiara rappresentazione del funzionamento del modello. Accettati i valori dell'aspettativa di vita alla nascita come input, esso fornisce la superficie della mortalità. Lo stesso approccio può essere utilizzato per qualsiasi altra misura demografica, come ad esempio la distribuzione dell'età alla morte.



Rappresentazione grafica del funzionamento del modello

Nella costruzione di una Neural Network, particolare importanza è prestata per la calibrazione (o fase di training o di learning). L'incertezza del modello è data dalla presenza dei pesi, che possono essere intesi come parametri incogniti: sono essi l'oggetto del training. Per determinarli, è necessario disporre di un certo numero di dati sperimentali relativi al parametro (o ai parametri) che la rete deve predire, sulla base dei quali è possibile "fittare" gli output predetti dalla rete stessa. L'operazione di fitting degli output consiste nell'aggiustamento dei pesi e quindi nel training.

Una volta definita la forma dell'output ed il contenuto del dataset da cui estrarre le variabili di input, si suddivide la serie storica in sottoperiodi, i quali delimitano l'ambito di apprendimento (training set) e di valutazione. Quest'ultimo viene in genere distinto in test set e generalisation set. Alla fase di training, volta all'apprendimento della rete, segue quella di testing per

il completamento della modellazione della NN. Questa consiste nella valutazione delle prestazioni del modello costruito mediante l'utilizzo di un testing set di dati, che deve essere diverso dal training set usato precedentemente. Nello specifico, le analisi effettuate considerano per entrambe le fasi i dati relativi agli anni che vanno dal 1950 al 2005.

La procedura di testing, per verificare l'applicabilità della rete neurale, consiste nella suddivisione del dataset in due parti: l'80% dei dati viene impiegata nelle procedure di learning e training, mentre il restante 20% viene utilizzato per avviare la simulazione della rete e controllare i risultati attraverso il testing.

In questo contesto la procedura di training risulta leggermente diversa dall'esercizio canonico di backtesting, in particolare il vettore considerato per il test viene scelto casualmente senza utilizzare una selezione dipendente dal tempo. A titolo esemplificativo, nel vettore di addestramento (training vector) è possibile ottenere i seguenti anni: 1980, 1995, 1970, 2000. Al contrario, nel vettore di prova (test vector), possiamo considerare: 2004, 1985, 1999, 1997, 1975, 2001. I vettori del training test dunque, non seguono un modello time-dependent.

La capacità di generalizzazione della rete infine, è osservata mediante la fase di validation. A tal proposito, si considerano i dati relativi agli anni che vanno dal 2006 fino al 2014: la bontà previsionale del modello viene dunque valutata su questi 9 anni, calcolando lo scostamento ottenuto, ad ogni età, rispetto ai tassi effettivamente descritti nello Human Mortality Data Base. L'accuratezza dei risultati per ciascun paese, è verificata ulteriormente dal confronto delle rispettive performance con quelle derivanti dal modello Lee-Carter. Si utilizza una estensione del modello, la quale ipotizza che i decessi siano distribuiti come una Poisson: rispetto al modello originario tiene conto della diversa variabilità del fenomeno alle varie età.

La rete dunque, impara cercando di riconoscere la dinamica del training set, verifica come si adatta sul test set, e si applica infine, su un insieme di dati (validation) che non ha mai potuto osservare.

Particolare attenzione è stata prestata alla costruzione del modello da utilizzare per la proiezione della superficie di mortalità. A tal fine è stata effettuata una fase preliminare di ottimizzazione, volta a selezionare la miglior combinazione di iperparametri nella struttura della rete. Il numero ottimale di epoche e di neuroni per i diversi strati è stato stabilito sottoponendo più volte al training la rete, variando opportunamente il numero di neuroni o di epoche ed annotando l'accuratezza della risposta, in termini di errori, sul

testing set. Sulla base di tali esperimenti, viene scelto di volta in volta, per ciascun paese e genere, un determinato numero di unità e di epoche.

Con riferimento alla numerosità dei neuroni che devono essere assegnati a ciascuno strato nascosto, il criterio utilizzato nella scelta è quello della minimizzazione del rischio di overlearning che si corre quando si decide di inserire un numero eccessivo di neuroni. In questo caso, le unità disegnano quasi perfettamente il pattern della serie storica ma non sono in grado di generare una previsione affidabile poichè riducono il contributo degli input. Viceversa, il rischio di assegnare un numero troppo basso di neuroni è quello di ridurre il potenziale di apprendimento della rete neurale. In particolare, la fase di ottimizzazione ha portato a scegliere, per tutti i paesi studiati, architetture costituite da sei hidden layers. Per ciascun layer, viene utilizzato un numero diverso di neuroni in grado di produrre risultati coerenti con la realtà.

La funzione di attivazione utilizzata è la Rectified Linear Unit (ReLU), la migliore tra le altre testate. Si tratta di una funzione molto semplice da calcolare: appiattisce a zero la risposta a tutti i valori negativi, mentre lascia tutto invariato per valori uguali o superiori a zero. Questa caratteristica la rende una funzione particolarmente utile nei layers intermedi, dove la quantità di passaggi e di calcoli è notevole. Calcolare la derivata infatti, risulta così molto semplice: per tutti i valori negativi è pari a zero, mentre per quelli positivi è uguale ad 1. Nel punto angoloso nell'origine della funzione invece, la derivata è indefinita ma viene comunque posta nulla, per convenzione.

Durante la procedura di fitting, vengono eseguite 150 epoche per la maggior parte dei paesi studiati, con batch size pari ad uno. Non vi sono infine altri iperparametri da segnalare, influenti sulla performance.

4.3 Risultati

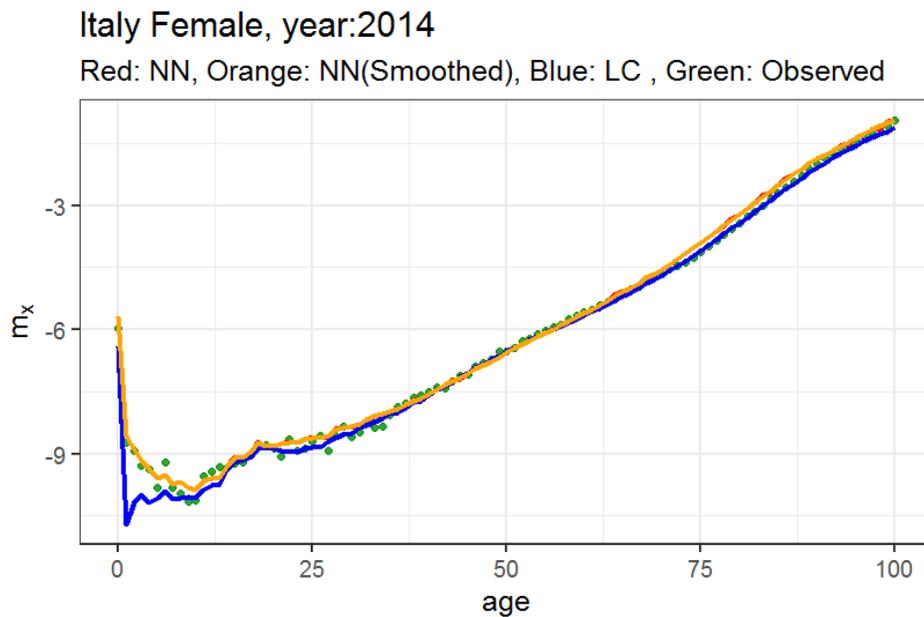
4.3.1 Italia

La rete neurale profonda utilizzata per il genere maschile e femminile, è composta in entrambi i casi da:

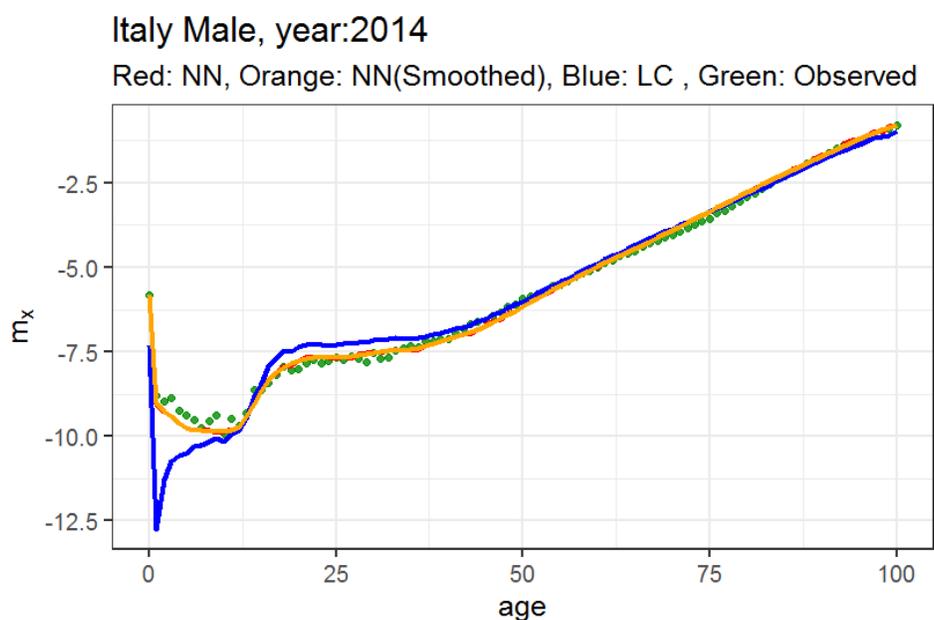
- 1 input layer composto da 600 neuroni,
- 6 hidden layers composti da 600 neuroni ciascuno,

- 1 output layer composto da 101 neuroni, corrispondenti ai tassi di mortalità per le età da 0 a 100.
- 150 epoche per l'addestramento.
- funzione di attivazione ReLU.

I seguenti grafici mostrano l'andamento nel tempo dei tassi di mortalità femminili e maschili in Italia nel 2014. Più in dettaglio, i tassi di mortalità osservati, nel grafico in verde, sono confrontati con quelli ottenuti dal modello Lee-Carter (in blu) e quelli derivanti dal modello di rete neurale profonda (in rosso o arancione, se si tiene conto dello smoothing¹).

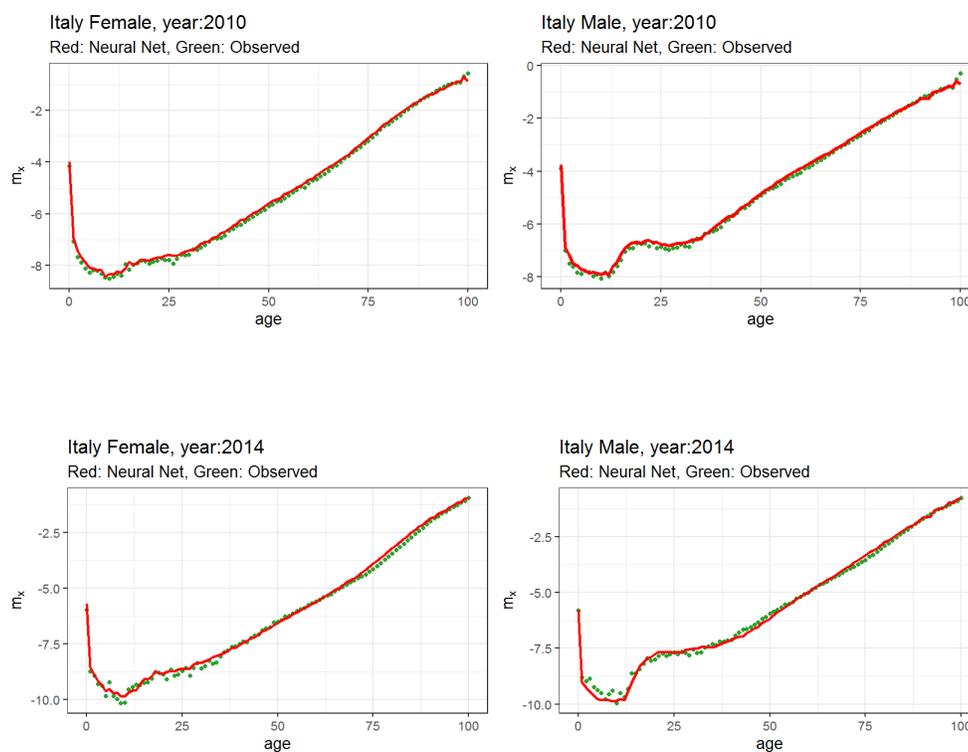


¹Consiste nell'applicazione di una funzione di filtro il cui scopo è evidenziare i pattern significativi tra i dati, attenuando il rumore generato da fenomeni di disturbo



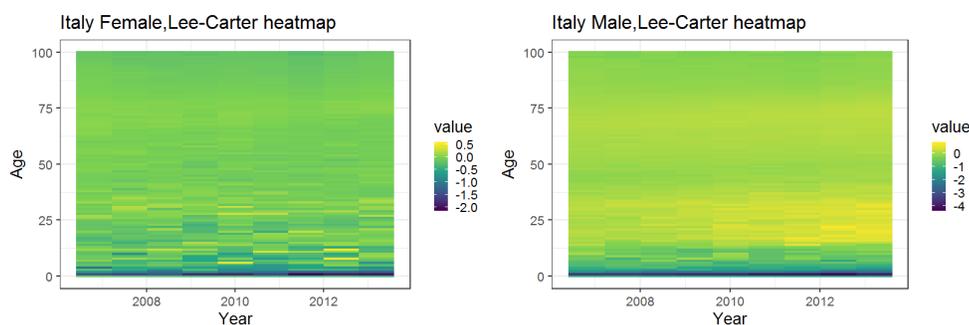
Come si evince dai grafici, il modello di rete neurale garantisce una buona stima dei tassi di mortalità, evidente dalla ridotta distanza tra i punti in verde e la curva arancione. Rispetto al Lee-Carter, offre inoltre una maggiore aderenza alla realtà, specialmente nei primi anni di vita.

I buoni risultati derivanti dall'utilizzo del modello di rete neurale sono evidenti anche dall'osservazione dei grafici dell'andamento della mortalità femminile e maschile, in Italia, negli anni che vanno dal 2006 al 2014. Verranno ora mostrati quelli ottenuti nel 2010 e nel 2014, confrontati questa volta solo con la mortalità osservata.



Nuovamente, l'osservazione dei grafici ci permette di affermare che il modello di rete neurale profonda utilizzato fornisce dei risultati coerenti con quanto descritto nello Human Mortality Database. I tassi stimati infatti, sopra rappresentati in rosso, mostrano uno scostamento da quelli osservati (in verde) decisamente ridotto, sintomo quindi della bontà del meccanismo di proiezione su cui si basa il modello.

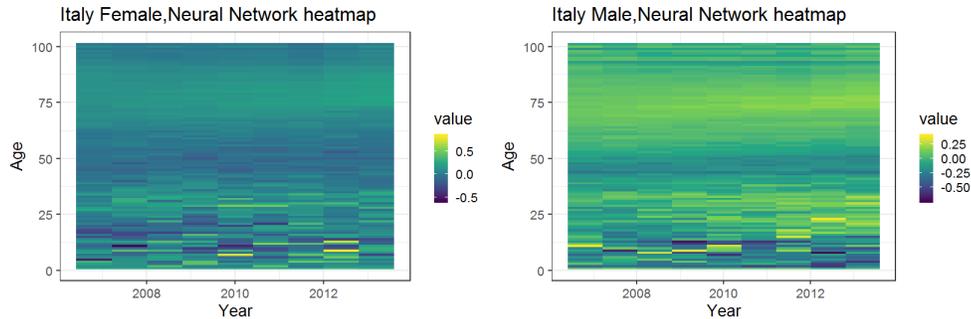
La capacità previsionale della rete è suggerita anche dall'osservazione delle heatmaps relative alla differenza tra i tassi di mortalità predetti e quelli reali. Nello specifico, nei seguenti grafici si confronta lo scarto di tasso per età e anno di calendario ottenuto a partire dal modello Lee-Carter e quello prodotto dal modello di rete neurale, mantenendo la distinzione per sesso.



I grafici fanno riferimento al modello Lee-Carter. La scala dei colori indica l'intensità della differenza di tasso riscontrata. Nella popolazione femminile lo scostamento (negativo) maggiore si registra nei primi anni di età, per tutti gli anni di calendario. Qui infatti, il modello tende a sottostimare il fenomeno di mortalità su tutto l'orizzonte temporale considerato. Lo scarto risulta poi annullarsi al crescere dell'età, sebbene circa fino ai 60 anni si evidenziano alcuni lievi scostamenti per ogni anno di calendario. Superati i 60 anni la differenza di tasso risulta omogeneamente prossima allo zero, come si evince dai segmenti verdi.

Situazione analoga si riscontra nel caso maschile. Lo scarto risulta di nuovo maggiore nei primi anni di vita, per ogni anno di calendario, per poi vedere una progressiva diminuzione al crescere di essa. Superati i 40 anni la differenza di tasso è tendenzialmente prossima allo zero, ogni anno. Da segnalare è invece quanto accade tra i 10 ed i 30 anni. Come si evince dal grafico, in questa fascia il modello Lee-Carter tende a sovrastimare lievemente la mortalità al variare dell'anno di calendario, sebbene producendo uno scarto quasi nullo.

Il confronto dei due grafici suggerisce che per entrambi i sessi il modello fornisce una buona approssimazione dei tassi reali, tuttavia con un certo margine di errore per i primi anni di vita. L'intensità dello scarto risulta però differente nelle due popolazioni. Mentre tra le donne lo scostamento massimo risulta pari a -2, tra gli uomini questo risulta salire fino a -4, sinonimo quindi di una migliore bontà di rappresentazione del fenomeno da parte del modello per la popolazione femminile.



Concentrandoci invece sul modello di rete neurale, dalle seguenti heatmaps si evince una maggiore coerenza tra le previsioni e i dati osservati. In dettaglio, nel caso femminile si osserva che lo scostamento tra i tassi risulta decisamente prossimo allo zero, muovendosi nell'intervallo $[-1, 1]$. Mentre dalla nascita fino a circa 40 anni lo scostamento oscilla tra valori positivi e negativi, le successive fasce di età presentano degli andamenti più delineati. Nello specifico, tra i 40 e i 60 anni, il modello tende a sottostimare il fenomeno di mortalità in tutti gli anni oggetto di analisi, mentre dai 60 fornisce tassi lievemente superiori a quelli osservati, su tutto l'orizzonte temporale considerato.

Situazione analoga si presenta nel caso maschile. Fino ai 40 anni lo scostamento oscilla tra valori positivi e negativi, evidenziando però una crescente sovrastima dei tassi all'avanzare dell'anno di calendario. Tra i 40 ed i 50, il modello sottostima lievemente la mortalità per poi assestarsi intorno allo zero.

In un confronto con il Lee-Carter, sembra evidente che il modello di rete neurale profonda fornisce una previsione della mortalità più accurata per entrambi i sessi. Lo scostamento risulta non solo spesso prossimo allo zero, ma anche concentrato in un range decisamente più contenuto, definito dall'intervallo $[-1, 1]$.

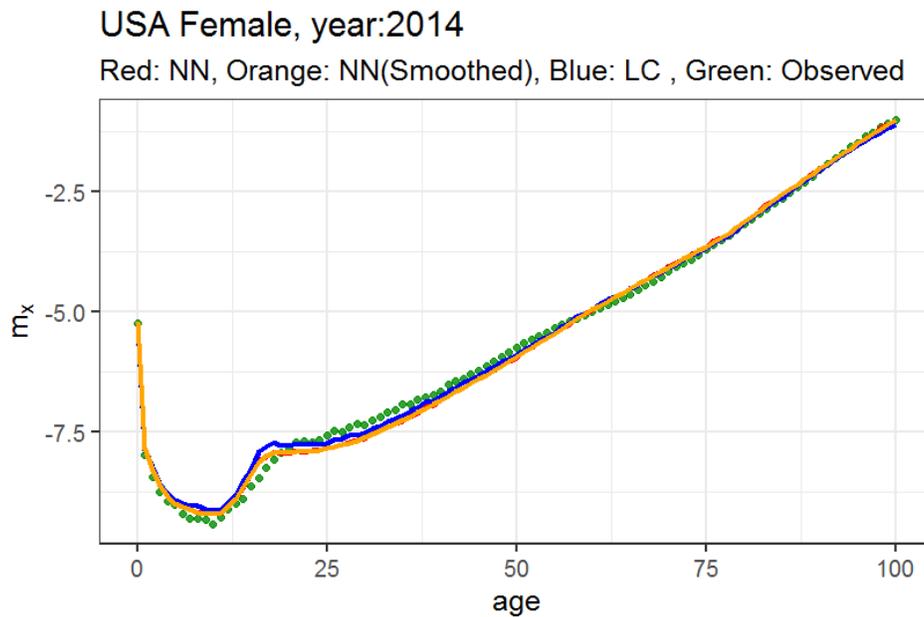
4.3.2 USA

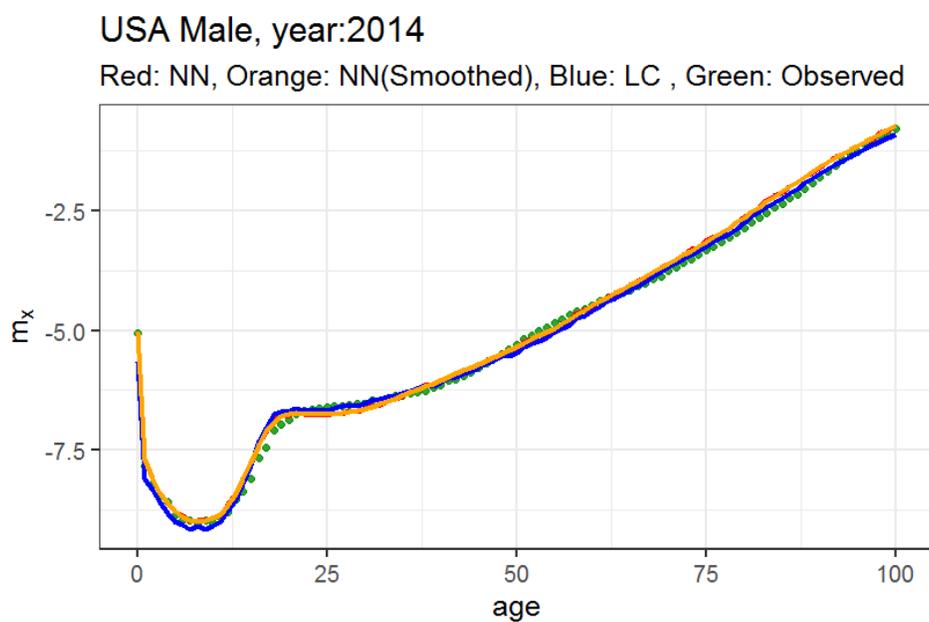
La rete neurale profonda utilizzata, è composta da:

- 1 input layer composto da 100 neuroni per il genere maschile e da 600 neuroni per il genere femminile.

- 6 hidden layers composti da 100 neuroni ciascuno per il genere maschile e da 600 neuroni ciascuno per il genere femminile,
- 1 output layer composto da 101 neuroni per entrambi i generi.
- 250 epoche per il genere maschile, 150 epoche per quello femminile.
- funzione di attivazione ReLU.

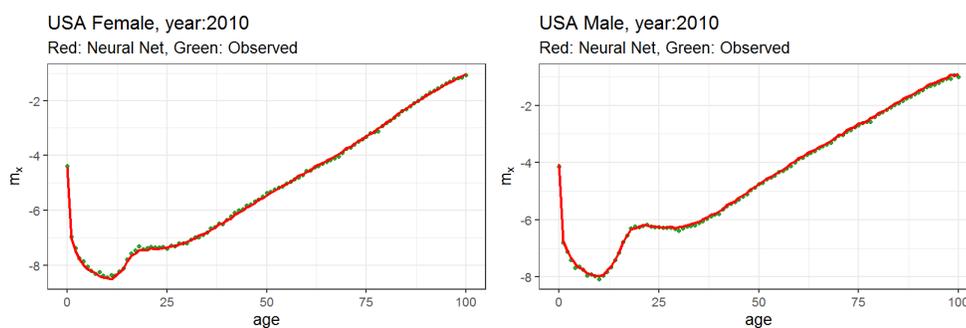
I grafici mostrano l'andamento nel tempo dei tassi di mortalità femminili e maschili in USA nel 2014. Nello specifico, i tassi di mortalità osservati, in verde, sono confrontati con quelli ottenuti dal modello Lee-Carter (in blu) e quelli derivanti dal modello di rete neurale profonda (in rosso o arancione, se si tiene conto dello smoothing).

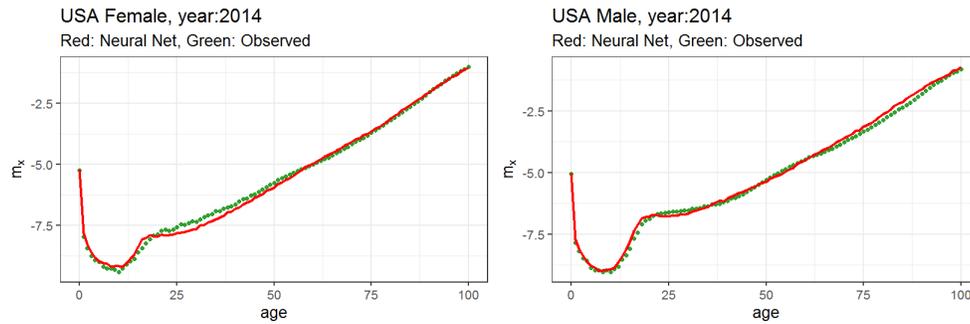




Come si evince dai grafici, il modello di rete neurale fornisce una buona stima dei tassi di mortalità, evidente dalla quasi sovrapposizione dei punti in verde con la curva arancione.

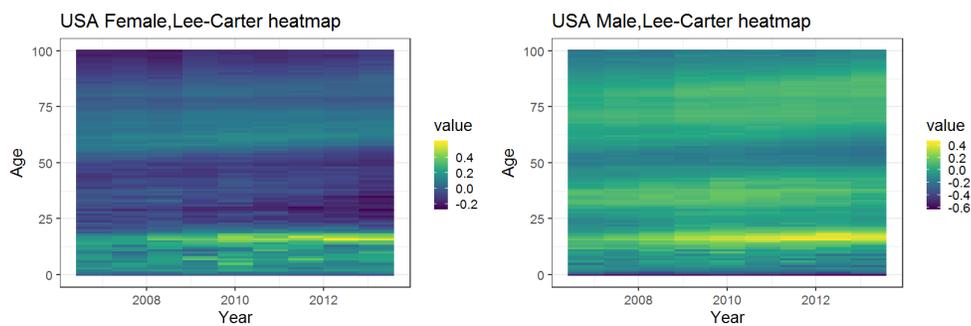
I buoni risultati derivanti dall'utilizzo del modello di rete neurale sono evidenti anche dall'osservazione dei grafici dell'andamento della mortalità femminile e maschile, negli USA, negli anni che vanno dal 2006 al 2014. Verranno ora mostrati quelli ottenuti nel 2010 e nel 2014, confrontati questa volta solo con la mortalità osservata.





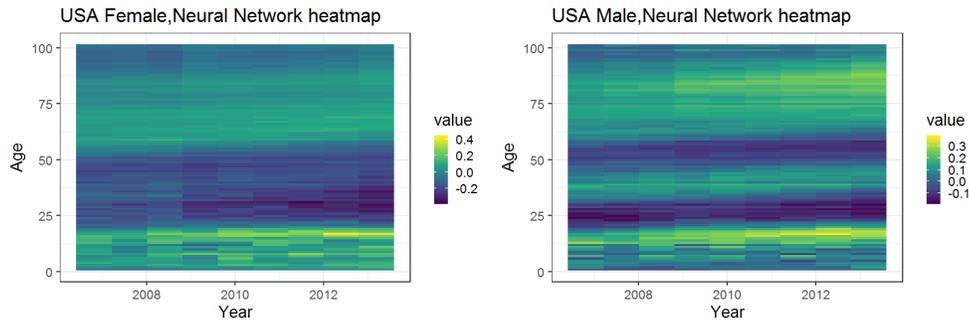
Nuovamente, l'osservazione dei grafici ci permette di affermare che il modello di rete neurale profonda utilizzato, fornisce dei risultati coerenti con quanto osservato. I tassi stimati infatti, sopra rappresentati in rosso, mostrano uno scostamento da quelli osservati (in verde) decisamente ridotto, sintomo quindi della bontà del meccanismo di proiezione su cui si basa il modello.

La bontà del modello è suggerita anche dall'osservazione delle heatmaps relative alla differenza tra i tassi di mortalità predetti e quelli reali. Nello specifico, nei seguenti grafici si confronta lo scarto di tasso per età e anno di calendario ottenuto a partire dal modello Lee-Carter e quello prodotto dal modello di rete neurale, mantenendo la distinzione per sesso.



Nella popolazione femminile il modello di previsione Lee-Carter fornisce un'ottima approssimazione dei tassi reali. Difatti, in modo sufficientemente omogeneo, per ogni anno di calendario e per ogni età, lo scostamento tra predizione ed osservazione è prossimo allo zero. Si registra esclusivamente una minima sottostima del fenomeno tra i 20 ed i 50 anni di età ed intorno ai 100, ed un lieve scarto positivo che cresce all'aumentare dell'anno di calendario, tra i 15 ed i 20 anni.

Situazione analoga si riscontra nel caso della popolazione maschile, in cui il modello Lee-Carter consente di ottenere ancora dei tassi di mortalità in linea con quanto descritto nello Human Mortality Database. E' possibile notare infatti che, alternando fasi di lieve sottostima e sovrastima del fenomeno, la heatmap suggerisce scostamenti prossimi allo zero per ogni età e per ogni anno di calendario. E' tuttavia interessante segnalare quanto accade su tutto l'orizzonte temporale tra i 15 ed i 20 anni: in questa fascia, all'aumentare dell'anno di calendario, aumenta lo scarto tra predizione ed osservazione, segno quindi di una lieve sovrastima del fenomeno nell'intervallo considerato.



L'utilizzo del modello di rete neurale, come si evince dalle heatmaps, comporta un'ulteriore, seppur parziale, miglioramento della predizione. Nel caso della popolazione femminile, si riscontra un caso simile a quanto descritto per il Lee-Carter. Per ogni età e anno di calendario, si riduce lo scostamento tra predizione ed osservazione. Nuovamente, si alternano fasi di sovrastima e

sottostima del fenomeno di mortalità. Da segnalare, nei primi anni di età, vi è uno scostamento positivo crescente all'aumentare dell'anno di calendario. Tra i 20 ed i 50 anni, si osserva un lieve scostamento negativo, indice di una sottostima del fenomeno oggetto di studio. Infine, oltre i 50 anni, si nota una progressiva riduzione dello scarto e quindi l'avvicinamento al valore zero.

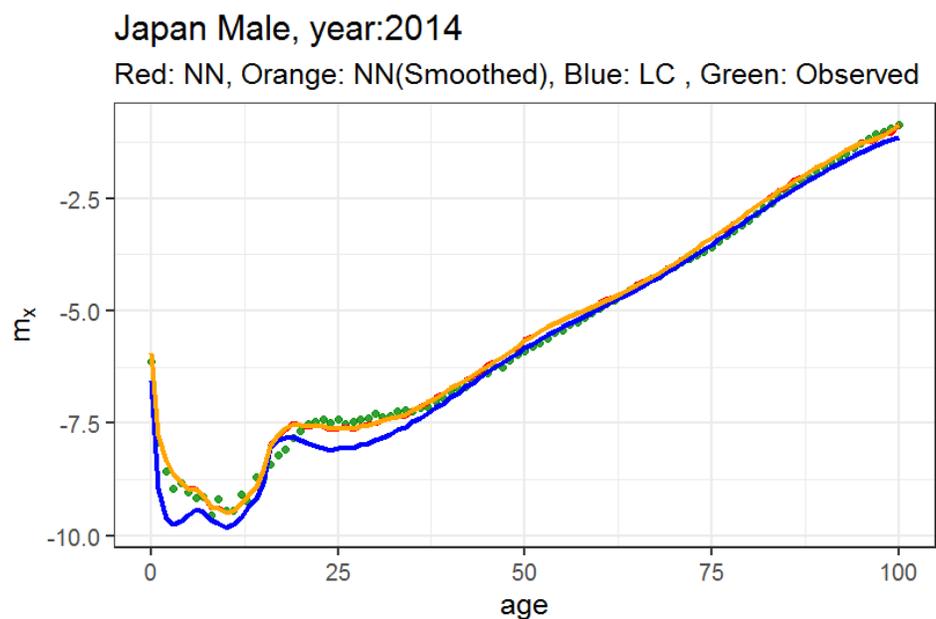
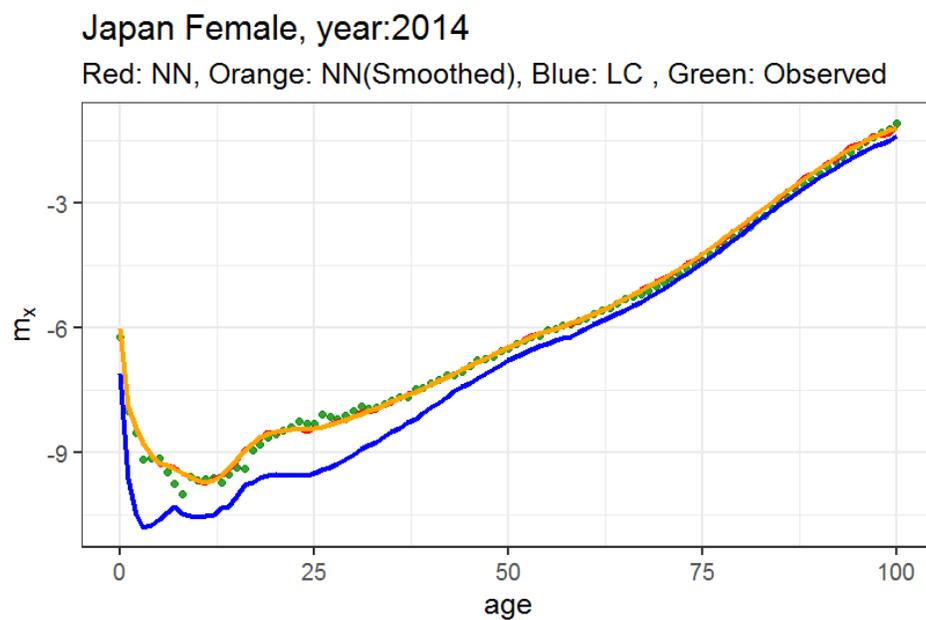
Nel caso della popolazione maschile, gli scostamenti tra predizione ed osservazione tendono nuovamente a diminuire, riducendo l'intervallo su cui varia l'intensità dello scarto. Si osserva l'alternarsi di fasi di sottostima e lieve sovrastima del fenomeno di mortalità, con valori tendenzialmente prossimi allo zero. Si segnala inoltre che, tra i 15 ed i 20 anni, lo scarto tende ad incrementarsi positivamente all'aumentare dell'anno di calendario.

4.3.3 Giappone

La rete neurale profonda utilizzata, è composta da:

- 1 input layer composto da 300 neuroni per il genere maschile e 600 neuroni per il genere femminile,
- 6 hidden layers composti da 300 neuroni ciascuno per il genere maschile e da 600 neuroni ciascuno per il genere femminile,
- 1 output layer composto da 101 neuroni in entrambi i casi,
- 150 epoche sia per il genere maschile che femminile,
- funzione di attivazione ReLU.

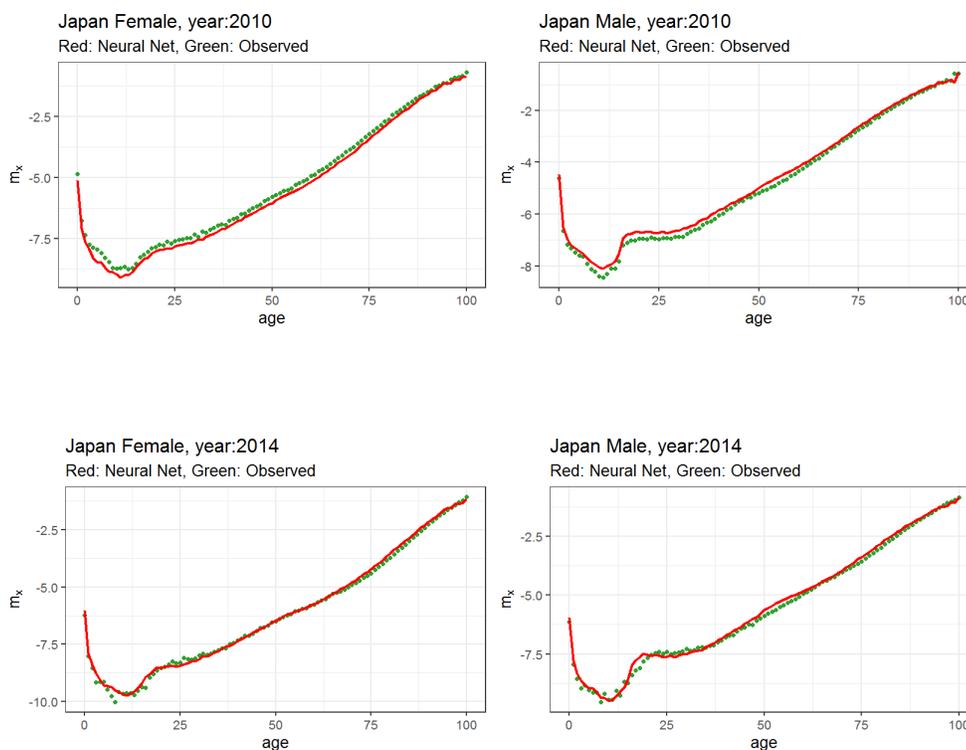
I grafici mostrano l'andamento nel tempo dei tassi di mortalità femminili e maschili in Giappone nel 2014. Nello specifico, i tassi di mortalità osservati, in verde, sono confrontati con quelli ottenuti dal modello Lee-Carter (in blu) e quelli derivanti dal modello di rete neurale profonda (in rosso o arancione, se si tiene conto dello smoothing).



Come si evince dai grafici, il modello di rete neurale garantisce un'ottima stima dei tassi di mortalità, evidente dalla ridotta distanza tra i punti in

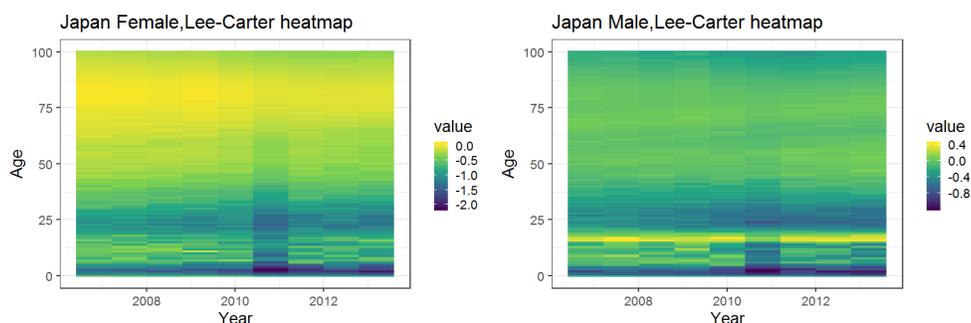
verde e la curva arancione. Rispetto al Lee-Carter, offre inoltre una maggiore aderenza alla realtà, specialmente nei primi 50 anni di vita circa.

I buoni risultati derivanti dall'utilizzo del modello di rete neurale sono evidenti anche dall'osservazione dei grafici dell'andamento della mortalità femminile e maschile, in Giappone, negli anni che vanno dal 2006 al 2014. Verranno ora mostrati quelli ottenuti nel 2010 e nel 2014, confrontati questa volta solo con la mortalità osservata.



Constatando quanto detto, l'osservazione dei grafici ci permette di affermare che, il modello di rete neurale profonda utilizzato fornisce dei risultati coerenti con quanto descritto nello Human Mortality Database. I tassi stimati infatti, sopra rappresentati in rosso, mostrano uno scostamento da quelli osservati (in verde) decisamente ridotto, sintomo quindi della bontà del meccanismo di proiezione su cui si basa il modello.

L'accuratezza delle stime ottenute dal modello è suggerita anche dall'osservazione delle heatmaps relative alla differenza tra i tassi di mortalità predetti e quelli reali. Nello specifico, nei seguenti grafici si confronta l'evoluzione nel tempo dello scarto di tasso per età e per anno di calendario ottenuto a partire dal modello Lee-Carter e quello prodotto dal modello di rete neurale, mantenendo la distinzione per sesso.

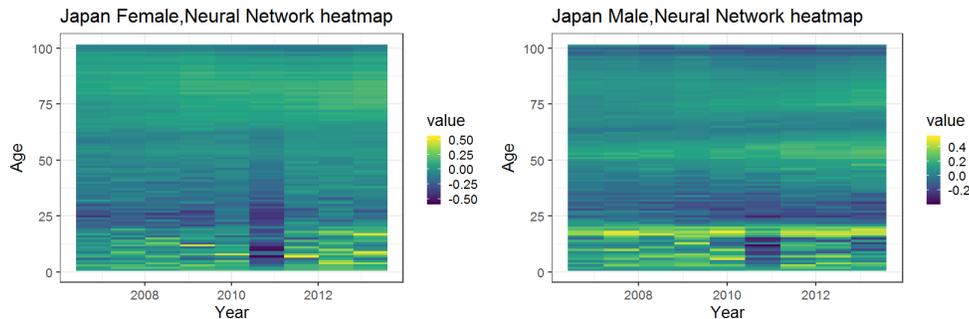


Nel primo grafico, relativo alla popolazione femminile secondo il modello Lee-Carter, lo scostamento tra i tassi predetti e quelli osservati ha il suo apice nei primi anni di età. Al crescere dell'età, omogeneamente su tutti gli anni di calendario, lo scarto tra predizioni e osservazioni tende a diminuire ed infine ad assestarsi, specificatamente dopo i 60 anni, attorno allo zero. Osservando il grafico si segnala dunque una sottostima del fenomeno per gran parte delle età.

Caso analogo si riscontra per la popolazione maschile. Il differenziale di tasso raggiunge i suoi valori più alti nei primi 20 anni, per poi diminuire ed assestarsi nuovamente attorno allo zero, al crescere dell'età della popolazione di riferimento. E' infine interessante segnalare quanto accade alle età di 15 e 16 anni: lo scostamento è sempre superiore, sintomo di una costante sovrastima, per ogni anno di calendario, della mortalità prevista col Lee-Carter rispetto a quella effettivamente osservata.

Il confronto tra le heatmaps relative alla popolazione femminile e maschile, consente di osservare come la previsione ottenuta utilizzando il modello

Lee-Carter nella popolazione Male sia più accurata di quella ottenuta per la popolazione Female. Infatti, lo scostamento massimo in quest'ultima può raggiungere valori pari a -2 , mentre nel caso Male l'intervallo di valori che lo scostamento assume è $[-1, 0.4]$.



Si osservano infine i grafici relativi allo scostamento tra i tassi predetti e quelli osservati in termini di rete neurale. Si evince che, nella popolazione femminile, l'utilizzo di una rete neurale comporta una sostanziale diminuzione dello scostamento tra predizione e osservazione, sintomo di una stima dei tassi più accurata rispetto a quella ottenuta utilizzando il metodo tradizionale. Si nota infatti che, in modo sufficientemente omogeneo, i tassi forniti dal modello si differenziano in maniera decisamente ridotta da quelli realmente osservati.

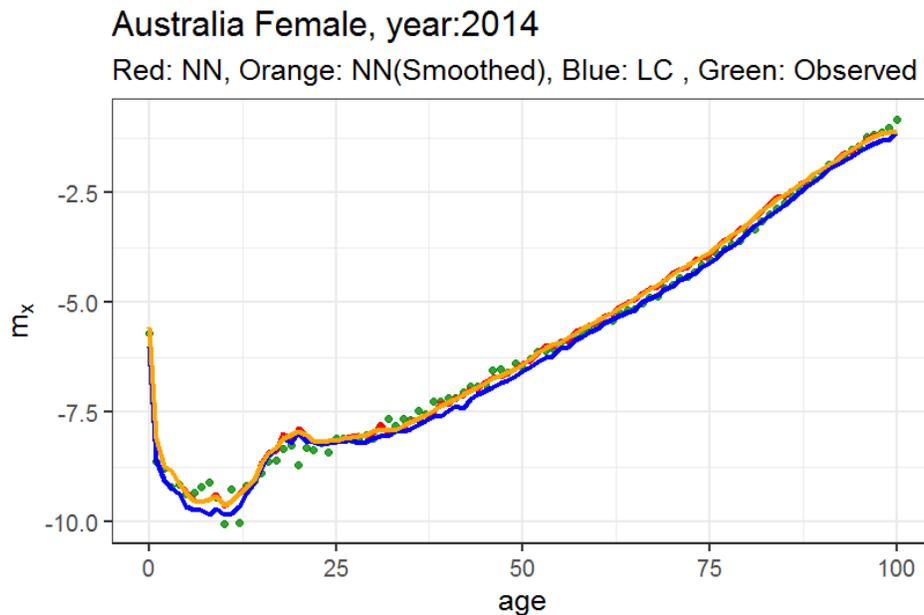
Caso analogo si riscontra nel caso maschile. I tassi predetti risultano decisamente prossimi a quelli reali, come è evidente anche dall'intervallo di valori $[-0.4, 0.4]$ in cui varia lo scostamento. Viene tuttavia fatta eccezione per quelli relativi ai primi anni di età: fino ai 20 anni circa, il modello sembra, per diversi anni di calendario, lievemente sovrastimare il fenomeno di mortalità rispetto ai tassi reali. Superati i 20 anni però, la predizione tende a coincidere sempre di più con l'osservazione.

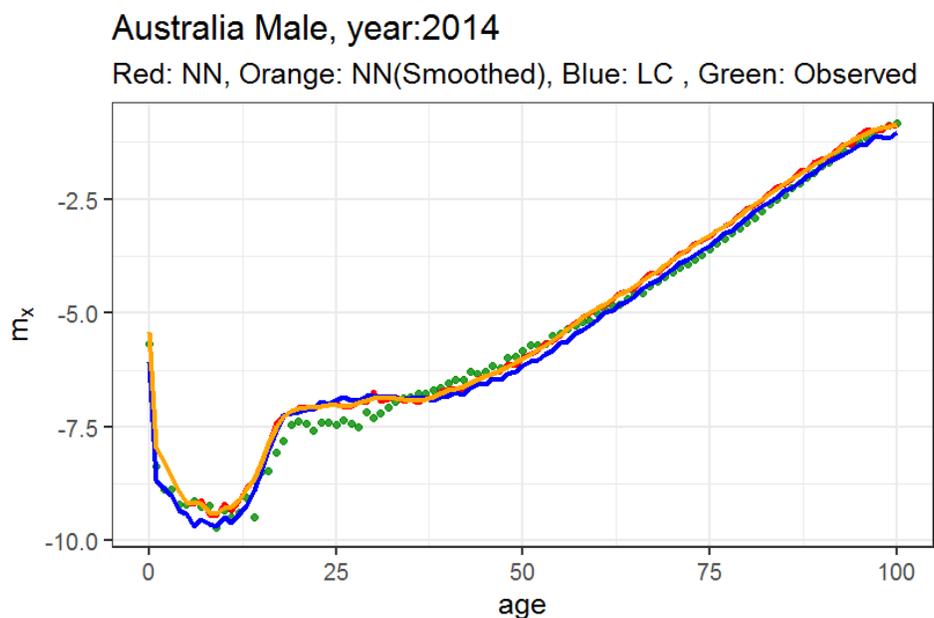
4.3.4 Australia

La rete neurale profonda utilizzata per il genere maschile e femminile, è composta in entrambi i casi da:

- 1 input layer composto da 600 neuroni,
- 6 hidden layers composti da 600 neuroni ciascuno,
- 1 output layer composto da 101 neuroni, corrispondenti ai tassi di mortalità per le età da 0 a 100.
- 150 epoche per l'addestramento.
- funzione di attivazione ReLU.

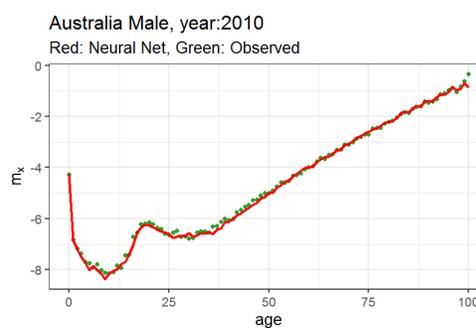
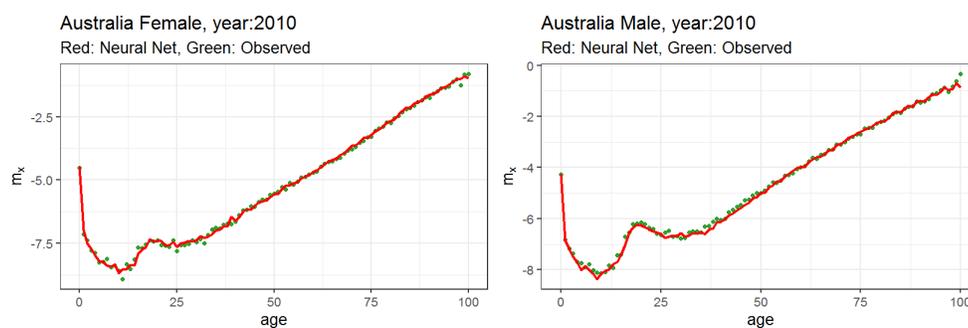
I grafici mostrano l'andamento nel tempo dei tassi di mortalità femminili e maschili in Australia nel 2014. Nello specifico, i tassi di mortalità osservati, in verde, sono confrontati con quelli ottenuti dal modello Lee-Carter (in blu) e quelli derivanti dal modello di rete neurale profonda (in rosso o arancione, se si tiene conto dello smoothing).

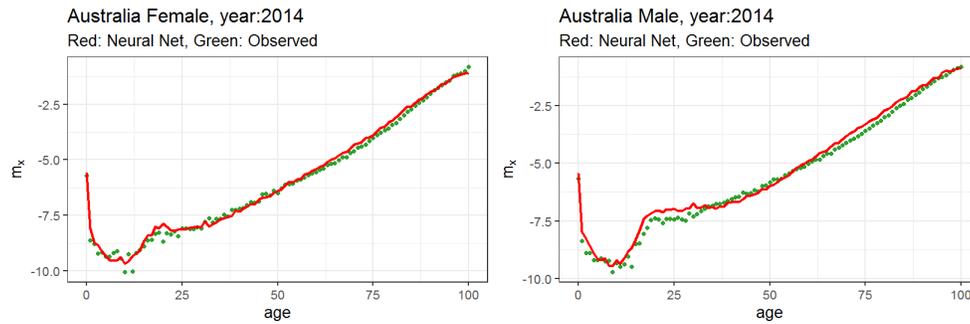




Come si evince dai grafici, il modello di rete neurale garantisce una buona stima dei tassi di mortalità, evidente dalla ridotta distanza tra i punti in verde e la curva arancione.

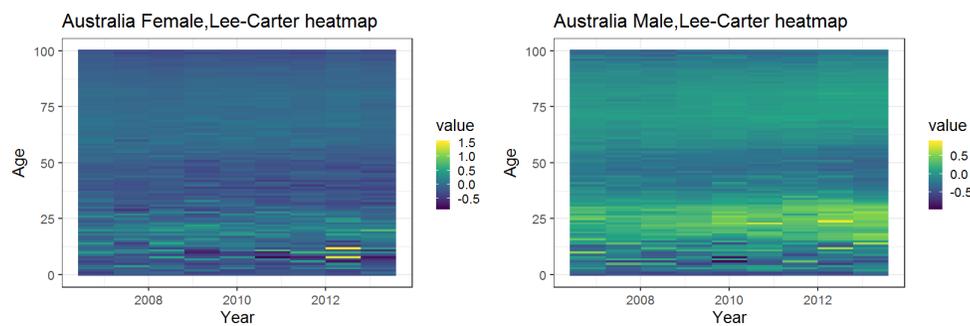
I buoni risultati derivanti dall'utilizzo del modello di rete neurale sono evidenti anche dall'osservazione dei grafici dell'andamento della mortalità femminile e maschile, in Australia, negli anni che vanno dal 2006 al 2014. Verranno ora mostrati quelli ottenuti nel 2010 e nel 2014, confrontati questa volta solo con la mortalità osservata.





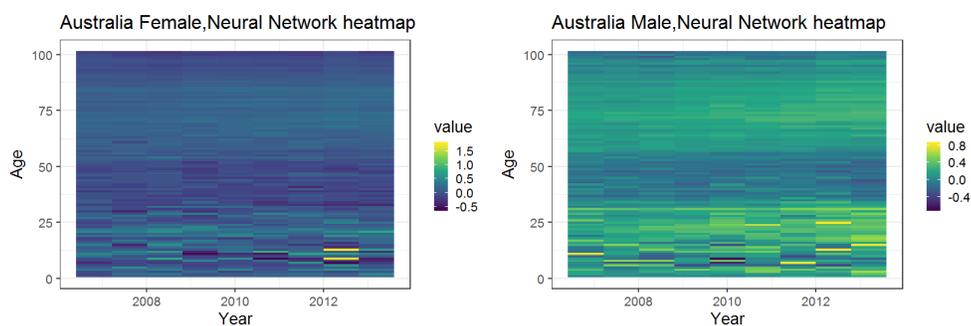
Nuovamente, l'osservazione dei grafici ci permette di affermare che il modello di rete neurale profonda utilizzato fornisce dei risultati coerenti con quanto descritto nello Human Mortality Database. I tassi stimati infatti, sopra rappresentati in rosso, mostrano uno scostamento da quelli osservati (in verde) decisamente ridotto, sintomo quindi della bontà del meccanismo di proiezione su cui si basa il modello.

La bontà del modello è suggerita anche dall'osservazione delle heatmaps relative alla differenza tra i tassi di mortalità predetti e quelli reali. Nello specifico, nei seguenti grafici viene descritto l'andamento nel tempo dello scostamento tra tassi predetti ed osservati, descriminando per sesso e per età, sia in termini di Lee-Carter che di rete neurale profonda.



Nel caso femminile, il modello Lee-Carter fornisce dei tassi predetti prossimi a quelli osservati, definendo quindi uno scarto molto vicino allo zero. Questa valutazione è omogenea per ogni anno di calendario e per ogni età considerata. L'osservazione della heatmap suggerisce inoltre, come il modello tenda lievemente a sottostimare il fenomeno di mortalità, eccetto nei primi anni di età, tra i 0 ed i 25 anni nello specifico, in cui si registrano anche delle situazioni di sovrastima dei tassi, come si evince dai segmenti gialli in corrispondenza dell'anno 2013.

Per quanto riguarda la popolazione maschile, si osserva nuovamente una buona approssimazione dei tassi predetti rispetto a quelli osservati. L'analisi del grafico suggerisce infatti come lo scarto tra predizione ed osservazione sia prossimo allo zero per ogni anno di calendario e per età. Fino ai 15 anni il modello tende a sottostimare parzialmente il fenomeno di mortalità, per poi mostrare una fase di sovrastima fino ai 30 anni circa. Tra i 30 ed i 60 anni, ed in prossimità dei 100, il LC tende a sottostimare ancora una volta il fenomeno oggetto di studio, per poi progressivamente realizzare uno scostamento rispetto ai tassi osservati decisamente prossimo allo zero, nella fascia di età che va dai 60 ai 90 anni.



Situazione analoga a quanto visto per il modello Lee-Carter, la rete neurale profonda per la popolazione femminile non fornisce un miglioramento sostanziale alla predizione dei tassi. Nuovamente, lo scostamento tra predizione ed osservazione risulta distribuito per età ed anno di calendario in

maniera simile a quanto già descritto per il modello LC. Tra i 0 ed i 25 anni, si registrano saltuariamente alcuni fenomeni di sovrastima della mortalità, legati alla presenza di segmenti gialli e verdi nella heatmap, mentre superati i 25 anni la colorazione blu suggerisce come il modello tenda a sottostimare i tassi di mortalità, nello specifico tra i 25 ed i 50 anni, ed in prossimità dei 100.

Nel caso della popolazione maschile, in maniera simile al caso precedente, la rete non sembra apportare migliorie significative in termini di qualità della predizione. Infatti, il range di valori osservati in termini di scarto, risulta essere maggiore rispetto a quello registrato utilizzando il modello Lee-Carter. Fino ai 30 anni circa, il modello tende a sovrastimare la mortalità rispetto a quanto descritto dallo Human Mortality Database. Tra i 30 ed i 60, si osserva una fase di sottostima della mortalità, a cui segue poi una fase in cui gli scostamenti risultano essere decisamente prossimi allo zero. Il confronto congiunto tra quanto ottenuto in termini di Lee-Carter e quanto ottenuto in termini di rete neurale profonda, suggerisce come quest'ultima non fornisca dei miglioramenti in termini di predizione rispetto al modello LC, pur mantenendo un'ottima bontà previsionale dei tassi di mortalità per gli anni considerati.

4.3.5 Russia

La rete neurale profonda utilizzata per il genere maschile e femminile, è composta in entrambi i casi da:

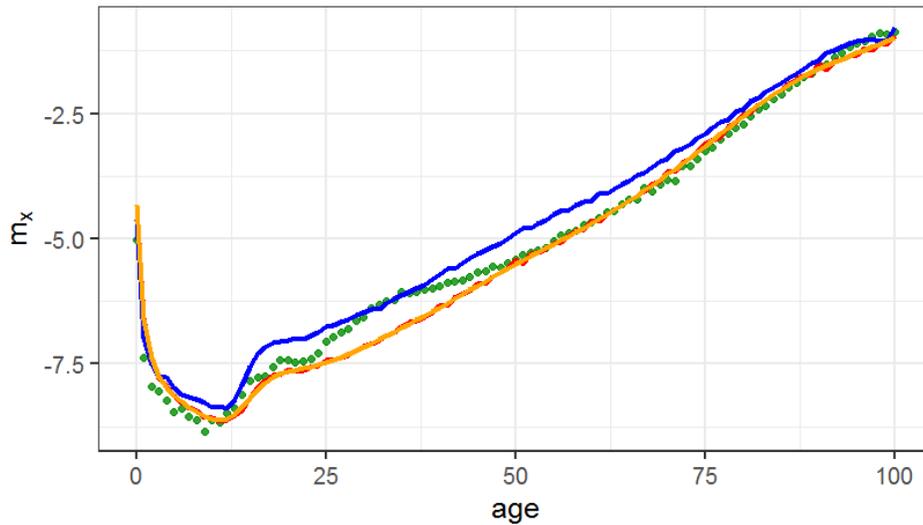
- 1 input layer composto da 100 neuroni,
- 6 hidden layers composti da 100 neuroni ciascuno,
- 1 output layer composto da 101 neuroni, corrispondenti ai tassi di mortalità per le età da 0 a 100.
- 150 epoche per l'addestramento.
- funzione di attivazione ReLU.

I grafici mostrano l'andamento nel tempo dei tassi di mortalità femminili e maschili in Russia nel 2014. Nello specifico, i tassi di mortalità osservati, in verde, sono confrontati con quelli ottenuti dal modello Lee-Carter (in blu) e

quelli derivanti dal modello di rete neurale profonda (in rosso o arancione, se si tiene conto dello smoothing).

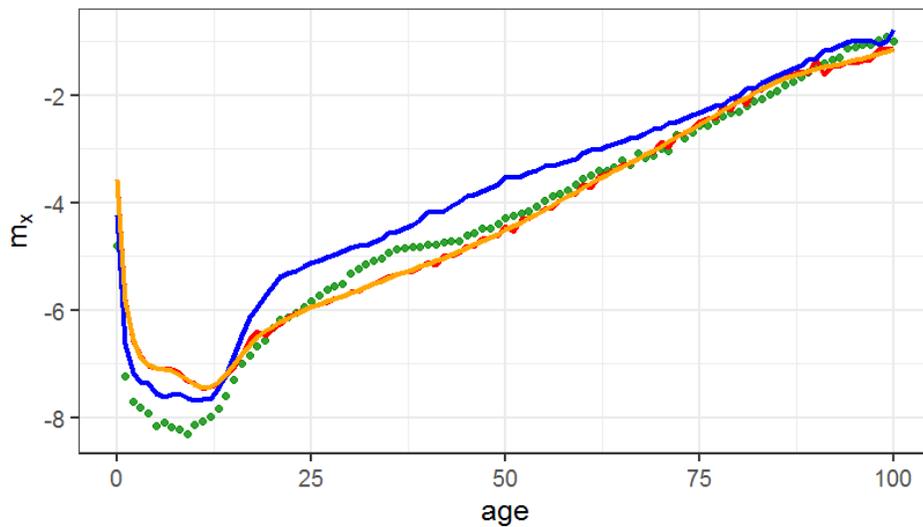
Russia Female, year:2014

Red: NN, Orange: NN(Smoothed), Blue: LC , Green: Observed



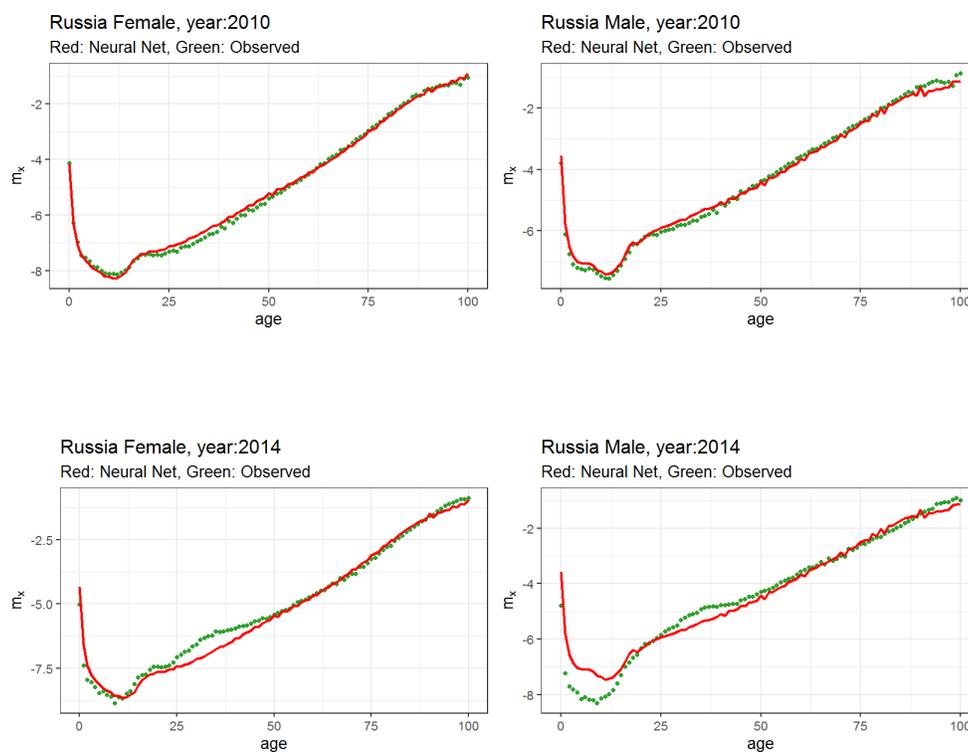
Russia Male, year:2014

Red: NN, Orange: NN(Smoothed), Blue: LC , Green: Observed



Come si evince dai grafici, per la Russia la proiezione dei tassi di mortalità risulta più complessa. Si osserva infatti che, sia il Lee-Carter che la rete neurale forniscono una stima ragionevole dei tassi di mortalità, seppur non completamente aderente alla realtà per parte delle età considerate.

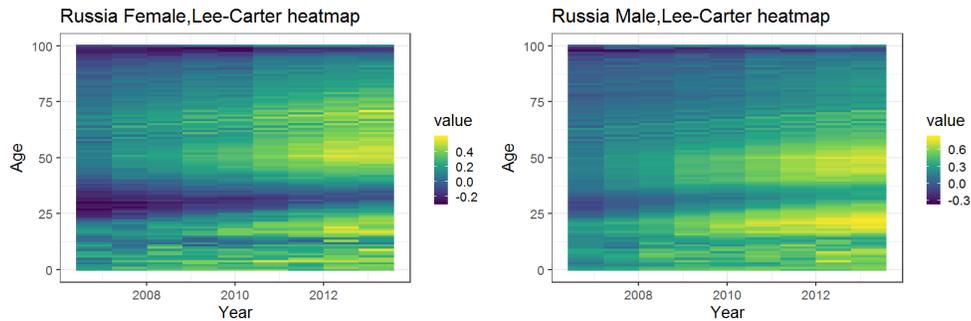
I risultati derivanti dall'utilizzo del modello di rete neurale sono evidenti anche dall'osservazione dei grafici dell'andamento della mortalità femminile e maschile, in Russia, negli anni che vanno dal 2006 al 2014. Verranno ora mostrati quelli ottenuti nel 2010 e nel 2014, confrontati questa volta solo con la mortalità osservata.



Nuovamente, l'osservazione dei grafici ci permette di affermare che il modello di rete neurale profonda utilizzato fornisce dei risultati tendenzialmente

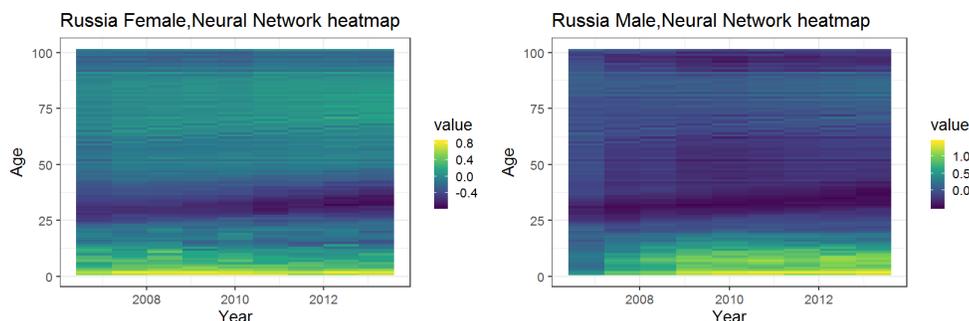
coerenti con quelli reali. I tassi stimati infatti, sopra rappresentati in rosso, mostrano uno scostamento ridotto da quelli osservati (in verde) nel 2010, maggiore invece nell'ultimo anno di forecasting.

Queste tendenze sono riscontrabili anche dall'analisi delle heatmaps.



In termini di Lee-Carter, nella popolazione femminile lo scostamento, inizialmente prossimo allo zero, tende ad aumentare al variare dell'anno di calendario, sovrastimando quindi il fenomeno di mortalità. Tra i 20 ed i 40 anni invece, lo scarto, in principio negativo, risulta progressivamente annullarsi di anno in anno. Questo meccanismo alternato di aumento/riduzione dello scostamento tende a replicarsi anche per le età successive.

Situazione analoga si riscontra nel caso maschile. Lo scarto risulta di nuovo esiguo nei primi anni di vita, per i primi anni di calendario, per poi vedere un progressivo incremento al variare dell'anno. Circa ogni 20 anni di vita, si replica il meccanismo alternato di aumento/riduzione dello scarto, già sperimentato tra la popolazione femminile.



A dispetto delle attese, il modello di rete neurale fornisce previsioni parzialmente meno coerenti con i dati osservati. In dettaglio, nel caso femminile si osserva che nei primi anni di età la rete sovrastima la mortalità per ogni anno di calendario, mentre fino a circa 50 anni il modello tende a sottostimare il fenomeno. Dai 50 anni in poi, lo scarto tra previsioni e dati reali risulta prossimo allo zero.

Il caso maschile presenta caratteristiche analoghe. Nei primi anni di età le stime sono superiori alle mortalità osservate, con uno scarto che cresce all'aumentare dell'anno di calendario. Nuovamente, tra i 20 e i 50 anni, le previsioni tendono a sottostimare la mortalità sull'intero orizzonte temporale considerato. Dopo i 50 anni circa, lo scostamento tra valori predetti ed osservati si avvicina consistentemente allo zero.

Un confronto congiunto tra le heatmaps ricavate dal modello Lee-Carter e quelle originate dalla rete neurale, suggerisce che gli scostamenti del primo sono più eterogenei ma di intensità minore.

4.3.6 Indici di bontà della stima

Una volta verificata la corretta dinamica temporale dell'errore, almeno in termini grafici, è necessario misurarla quantitativamente.

Tra i numerosi indicatori di errore sviluppati in ambito statistico, vengono considerati:

1. Il Mean Absolute Error (MAE),

$$MAE : \quad \sum^n \frac{|m_{x,t} - \hat{m}_{x,t}|}{n},$$

2. il Root Mean Square Error (RMSE),

$$RMSE : \sqrt{\frac{\sum^n (m_{x,t} - \hat{m}_{x,t})^2}{n}}$$

Si tratta di indicatori che misurano, in vario modo, il differenziale tra l'output desiderato e quello stimato dalla rete, in modo da valutare l'accuratezza previsionale della stessa. Solo nel caso in cui gli input, i neuroni, le funzioni di attivazione e tutti i parametri precedentemente descritti fossero perfettamente in grado di individuare il fenomeno oggetto di studio, gli scostamenti fra output reale e stimato sarebbero nulli, ottimizzando così i citati indicatori di errore. I buoni risultati forniti dalla rete neurale profonda sono infine evidenziati dal confronto delle sue performance, nella fase di validation, con quelle ottenute utilizzando il tradizionale modello Lee-Carter, per ciascun paese e genere.

<i>Country FEMALE</i>			<i>Country MALE</i>		
Italy	<i>MAE</i>	<i>RMSE</i>	Italy	<i>MAE</i>	<i>RMSE</i>
NN	0.1152	0.1519	NN	0.1137	0.1539
SmoothNN	0.1140	0.1506	SmoothNN	0.1097	0.1505
LC	0.1345	0.2630	LC	0.2680	0.5565
USA	<i>MAE</i>	<i>RMSE</i>	USA	<i>MAE</i>	<i>RMSE</i>
NN	0.1102	0.1338	NN	0.10494	0.1246
SmoothNN	0.1094	0.1325	SmoothNN	0.10390	0.1230
LC	0.1006	0.1306	LC	0.09787	0.1286
Japan	<i>MAE</i>	<i>RMSE</i>	Japan	<i>MAE</i>	<i>RMSE</i>
NN	0.10242	0.1353	NN	0.1133	0.1449
SmoothNN	0.09967	0.1332	SmoothNN	0.1123	0.1434
LC	0.45531	0.6216	LC	0.1793	0.2728
Australia	<i>MAE</i>	<i>RMSE</i>	Australia	<i>MAE</i>	<i>RMSE</i>
NN	0.1520	0.2101	NN	0.1770	0.2192
SmoothNN	0.1482	0.2055	SmoothNN	0.1732	0.2142
LC	0.1466	0.2043	LC	0.1719	0.2247
Russia	<i>MAE</i>	<i>RMSE</i>	Russia	<i>MAE</i>	<i>RMSE</i>
NN	0.2007	0.2717	NN	0.2589	0.3730
SmoothNN	0.1982	0.2693	SmoothNN	0.2532	0.3699
LC	0.2076	0.2477	LC	0.3057	0.3648

Dalle seguenti tabelle si evince che, a constatazione di quanto analizzato graficamente, le previsioni ottenute utilizzando la rete producono errori ridotti, prossimi e talvolta minori, a quelli del Lee-Carter. Anche nel caso della Russia, la cui complessità predittiva è riscontrata utilizzando lo stesso Lee-Carter, la rete neurale commette un errore circa pari a quello del modello di riferimento.

Capitolo 5

Conclusioni

L'elaborato propone un nuovo modello basato su una rete neurale feedforward e volto al forecasting dei tassi futuri di mortalità. Tra i vari metodi esistenti in letteratura, si fa specifico riferimento al modello Lee-Carter, di fatto il più noto e ampiamente utilizzato nella pratica demografica ed attuariale. Attraverso il confronto tra i due, si osserva la bontà previsionale della rete, capace di fornire stime talvolta più aderenti all'evidenza empirica.

L'andamento esponenziale del tasso centrale di mortalità, identificativo del modello tradizionale, ricalca perfettamente il fenomeno della decelerazione: la vita attesa per un soggetto di età x cresce nel tempo ma la velocità con cui ciò avviene tende a diminuire. Tale fenomeno tuttavia, accentuato nel corso degli ultimi 200 anni, ha portato ad un forte svantaggio per i modelli estrapolativi, ove l'indice di sensitività della mortalità per età, β_x , è caratterizzato da una struttura fissa che non permette di discernere in maniera precisa il fattore età rispetto al fattore tempo.

L'approccio proposto permette una notevolmente attenuazione del problema: indagata la struttura latente dell'aspettativa di vita, viene poi proiettata in conseguenza alla sua evoluzione. Lo stesso limite segnalato nel modello Lee-Carter, viene qui superato assumendo che i valori futuri della superficie di mortalità siano una funzione (stimata dalla Deep Neural Network) non lineare della proiezione dell'aspettativa di vita alla nascita.

Lo studio è stato condotto su cinque paesi, distinguendo per genere. La rete illustrata mostra, nella maggior parte dei casi, delle prestazioni molto accurate ed in linea con quanto osservato nello Human Mortality Database. Ciò è reso possibile grazie all'architettura profonda costruita per cogliere la relazione nascosta tra i dati.

In virtù dei risultati ottenuti, si conclude dunque che, il nuovo approccio proposto rappresenta un'utile alternativa ai metodi tradizionali utilizzati in letteratura per la proiezione dei tassi futuri. La stima della superficie di mortalità è ricavata mediante un solo dato in input, senza l'impiego di un'intera matrice di tassi e fornendo una previsione talvolta più accurata di quella ottenuta con lo stesso Lee-Carter.

Bibliografia

- [1] Alho, J. M. 1992. «Comment on «Modeling and Forecasting U.S. Mortality» by R. Lee and L. Carter.» *Journal of the American Statistical Association* 87(419, September):673–674.
- [2] Andrea Nigri, Susanna Levantesi, Mario Marino, Salvatore Scognamiglio and Francesca Perla (2019) «A Deep Learning Integrated Lee–Carter Model».
- [3] Andrès M. Villegas, Pietro Millosovich, Vladimir K. Kaishev. «StMoMo: An R Package for Stochastic Mortality Modeling»
- [4] Bell, W.R. 1997. «Comparing and Assessing Time Series Methods for Forecasting Age- Specific Fertility and Mortality Rates.» *Journal of Official Statistics* 13(3):279–303.
- [5] Bell, W.R. and B.C. Monsell. 1991. «Using Principal Components in time Series modeling and Forecasting of Age-Specific Mortality Rates.» *Proceedings of the American Statistical Association, Social Statistics Section* pp. 154–159.
- [6] Brouhns, N., M. Denuit and J. Vermunt. 2002. «A Poisson Log-bilinear Regression Approach to the Construction of Projected Lifetables.» *Insurance: Mathematics and Economics* 31:373–393.
- [7] Federico Girosi and Gary King (2007). « Understanding the Lee-Carter Mortality Forecasting Model»
- [8] Haberman S, Renshaw A (2009). «On Age-Period-Cohort Parametric Mortality Rate Projections.» *Insurance: Mathematics and Economics*, 45(2), 255–270.

- [9] King, Gary. 1989. *Unifying Political Methodology: The Likelihood Theory of Statistical Inference*. Michigan University Press.
- [10] Lee R, Miller T (2001). «Evaluating the Performance of the Lee-Carter Method for Forecasting Mortality.» *Demography*, 38(4), 537–549.
- [11] Lee, Ronald D. 2000a. «The Lee-Carter Method for Forecasting Mortality, with Various Extensions and Applications.» *North American Actuarial Journal* 4(1):80–93.
- [12] Lee, Ronald D. and Lawrence R. Carter. 1992. «Modeling and Forecasting U.S. Mortality.» *Journal of the American Statistical Association* 87(419, September).
- [13] Lee, Ronald D. and Timothy Miller. 2001. «Evaluating the Performance of the Lee-Carter Approach to Modeling and Forecasting Mortality.» *Demography* 38(4, November):537– 549.
- [14] McNown, Rober and Andrei Rogers. 1989. «Forecasting Mortality: A Parameterized Time Series Approach.» *Demography* 26(4):645–660.
- [15] McCullagh P, Nelder J (1989). «Generalized Linear Models.» second edition. Chapman & Hal, London.
- [16] Oeppen J, Vaupel JW (2002). «Broken Limits to Life Expectancy.» *Science*, 296(5570), 1029–1031.
- [17] Plat R (2009). «On Stochastic Mortality Modeling.» *Insurance: Mathematics and Economics*, 45(3), 393–404.
- [18] Renshaw A, Haberman S (2006). «A Cohort-Based Extension to the Lee-Carter Model for Mortality Reduction Factors.» *Insurance: Mathematics and Economics*, 38(3), 556–570.