



SAPIENZA
UNIVERSITÀ DI ROMA

Stima e proiezione della sopravvivenza: un focus sui percettori di rendita dell'INPS.

Facoltà di Ingegneria dell'Informazione, Informatica e Statistica
Corso di Laurea Magistrale in Scienze Attuariali e Finanziarie

Raffaello Iacono
Matricola 2038456

Relatore
Prof.ssa Susanna Levantesi

Correlatore
Prof. Fabio Baione

Anno Accademico 2023/2024

Indice

INTRODUZIONE	3
1 INTRODUZIONE ALLA MORTALITA'	6
1.1 Definizione e storia del pensiero sulla morte.....	6
1.2 La mortalità e i fattori influenti.....	6
1.3 Storia dello studio mortalità nel mondo.....	7
1.4 Elementi di studio della mortalità.....	9
1.5 La Tavola di Mortalità di una Generazione.....	11
2 IL LONGEVITY RISK	20
2.1 La Notazione.....	23
2.2 Modelli stocastici generalizzati di mortalità di tipo APC (Age-Period-Cohort).....	24
2.3 Il Modello Lee-Carter (1992). (M1).....	26
2.3.1 Stima dei parametri.....	29
2.3.2 Forecasting (proiezione dei parametri).....	32
2.3.3 Estensione di Brouhns et al. (2002).....	34
2.3.4 Estensione di Lee e Miller (2001).....	35
2.3.5 Estensione di Booth-Maindonald-Smith (BMS) (2002).....	37
2.4 Il modello di Renshaw-Haberman. (M2).....	38
2.5 Il modello di Cairns-Blake-Dowd (2006). (M5).....	38
2.5.1 Il modello Cairns-Blake-Dowd con Effetto Coorte. (M7).....	41
2.5.2 Il modello Cairns-Blake-Dowd con Effetto Coorte. (M6)-(M8).....	42
2.6 Il modello Plat.....	42
2.7 I criteri di scelta del modello.....	43
3 INTELLIGENZA ARTIFICIALE PER LA MODELLAZIONE E PREVISIONE DELLA MORTALITA'	44
3.1 Il Machine Learning.....	44
3.2 Machine e Deep Learning per lo studio della mortalità.....	46
3.2.1 Applicazione delle tecniche di Machine Learning.....	47
3.2.2 Applicazione delle tecniche di Deep Learning.....	59
4 ANALISI DELLA SOPRAVVIVENZA DEI PENSIONATI	69
4.1 Introduzione e obiettivi dell'analisi.....	69
4.1.1 Il sistema pensionistico in Italia.....	70
4.2 Introduzione ai dati.....	76
4.2.1 Descrizione del dataset.....	76
4.2.2 Pulizia del dataset.....	78

4.3	Statistiche descrittive preliminari.....	80
4.4	Life Expectancy e Life-span Inequality.....	93
4.4.1	Analisi Life Expectancy (e_x) e della Life-span Inequality (indice e -dagger $e \dagger$).....	94
4.4.2	Proiezione della Life Expectancy (e_x).	105
4.4.3	Proiezione dell'indice E-Dagger ($e \dagger$).	123
4.4.4	Confronto dei Modelli Stocastici di Mortalità: Analisi dei Valori BIC per le Diverse Combinazioni.	140
4.4.5	Analisi delle Traiettorie nelle Proiezioni.	142
	CONCLUSIONE.	149
	BIBLIOGRAFIA.	155

INTRODUZIONE.

Nell'ambito delle Scienze Attuariali, la mortalità e la sopravvivenza rappresentano due concetti centrali per l'analisi e la pianificazione di politiche previdenziali. Il tema delle pensioni coinvolge trasversalmente tutta la popolazione, rendendo essenziale la stima accurata della sopravvivenza per garantire la sostenibilità a lungo termine del sistema previdenziale. In Italia, questo tema è particolarmente rilevante poiché la maggior parte dei pensionati è iscritta a una delle gestioni dell'Istituto Nazionale della Previdenza Sociale (INPS), l'ente previdenziale più importante del Paese.

La mia tesi, intitolata "Stima e proiezione della sopravvivenza: un focus sui percettori di rendita dell'INPS", ha come obiettivo principale l'analisi e la proiezione delle dinamiche di sopravvivenza tra i percettori di pensione dell'INPS, mediante l'uso di modelli stocastici avanzati. Questa analisi permetterà di stimare due grandezze chiave: la Life Expectancy (aspettativa di vita) e l'indice E-Dagger che misura la disuguaglianza nella durata della vita. Questi due indicatori non solo sono fondamentali in campo statistico-attuariale, ma rivestono anche un ruolo cruciale per le decisioni politiche. Infatti, una stima accurata della sopravvivenza può supportare le istituzioni nella gestione del "Longevity Risk", ossia il rischio legato all'allungamento della vita media.

La scelta di approfondire questo tema è stata motivata non solo dalla sua rilevanza accademica, ma anche dal mio interesse personale per le conseguenze delle politiche previdenziali sulla vita delle persone. Garantire la sostenibilità dei sistemi pensionistici è una sfida che riguarda direttamente la sicurezza economica e il benessere della popolazione.

L'allungamento della vita e il conseguente invecchiamento della popolazione pongono sfide senza precedenti per i sistemi previdenziali. La gestione del longevity risk è infatti fondamentale per garantire la sostenibilità del sistema pensionistico a lungo termine, ed è su questo che si fonda la rilevanza di questo elaborato. Una corretta stima della sopravvivenza, infatti, non solo consente di valutare con precisione le risorse necessarie per il pagamento delle pensioni, ma permette anche di creare politiche previdenziali più eque e lungimiranti. Questo contribuisce a proteggere le generazioni future e a garantire la sostenibilità del sistema nel lungo periodo.

La tesi è organizzata in quattro capitoli, ognuno dei quali affronta un aspetto cruciale per l'analisi della mortalità e della sopravvivenza.

Il Capitolo 1, intitolato "Introduzione alla mortalità", introduce il concetto di morte e mortalità, fornendo un quadro storico del suo studio e descrivendo i principali strumenti utilizzati in ambito demografico. Il capitolo traccia un percorso storico che parte dal XVII secolo e arriva fino ai giorni nostri, mettendo in luce i fattori che influenzano la mortalità, come quelli socio-demografici, biologici e ambientali. In particolare, viene dato ampio

spazio alle tavole di mortalità, uno degli strumenti più importanti per l'analisi della sopravvivenza. Quindi, tale capitolo pone le basi teoriche necessarie per gli approfondimenti successivi.

Nel Capitolo 2, successivamente, il focus si sposta sul "Longevity Risk", spiegando come l'allungamento della vita media rappresenti un rischio significativo per le istituzioni che gestiscono fondi pensionistici e, inoltre, vengono introdotti i principali modelli stocastici di mortalità utilizzati in ambito attuariale per gestire e prevedere la mortalità futura. In particolare, si analizzano modelli come il Lee-Carter. Tale modello, sviluppato nel 1992, è uno dei modelli più noti per la previsione della mortalità e viene trattato in dettaglio insieme alle sue estensioni. Oltre ad esso, vengono analizzati anche altri modelli importanti, come ad esempio il modello Renshaw-Haberman, il modello Age-Period-Cohort e il modello Cairns-Blake-Dowd, più le relative estensioni. Ogni modello è descritto nel dettaglio, con una particolare attenzione al loro utilizzo per gestire il rischio di longevità e migliorare la precisione delle previsioni attuariali. L'obiettivo di questo capitolo è quindi, quello di spiegare la struttura di questi modelli e mostrare come essi possano essere utilizzati per prevedere l'evoluzione futura della mortalità e della sopravvivenza e, conseguentemente, supportare le decisioni previdenziali.

Nel Capitolo 3, si esplora l'utilizzo di tecniche di "Machine Learning" per migliorare ulteriormente la previsione della mortalità. Algoritmi come i Decision Trees, i Random Forest, il Gradient Boosting e le reti neurali sono introdotti come strumenti avanzati, complementari ai modelli stocastici tradizionali. Grazie alla loro capacità di identificare pattern complessi nei dati, questi algoritmi offrono previsioni più flessibili e precise, superando alcune delle limitazioni dei modelli lineari tradizionali. Questo capitolo, quindi, si focalizza sull'integrazione tra modelli attuariali e tecniche avanzate di intelligenza artificiale, spiegando come queste ultime possano essere utilizzate per affrontare sfide complesse come la previsione della mortalità, e conseguentemente della sopravvivenza, in scenari altamente incerti.

Il Capitolo 4, infine, si focalizza sull'analisi della mortalità e della sopravvivenza dei pensionati INPS, coprendo un periodo che va dal 1981 al 2016. L'obiettivo principale è stimare due indicatori chiave della sopravvivenza, la Life Expectancy (aspettativa di vita) e l'indice E-Dagger, per diverse sottopopolazioni, differenziate in base a sesso, categoria della pensione, gestione previdenziale e classe d'importo. L'analisi consente di evidenziare le disuguaglianze nella durata della vita tra gruppi di individui, fornendo una visione dettagliata delle differenze nella sopravvivenza.

Il primo passo dell'analisi ha riguardato la pulizia dei dati. Questo processo è stato necessario per eliminare eventuali anomalie e garantire che i dati fossero coerenti e utilizzabili per l'analisi statistica. Una volta completata la pulizia, i dati sono stati elaborati per evidenziare delle statistiche preliminari utili a fornire una prima idea di come gli

individui fossero suddivisi in base alle diverse variabili sopraelencate. Successivamente, sono state calcolate le due grandezze di sopravvivenza per ciascuna combinazione delle variabili rilevanti. Questi calcoli hanno permesso di individuare le principali tendenze nel tempo e di analizzare come la sopravvivenza sia cambiata tra il 1981 e il 2016.

A questo punto, l'analisi si è concentrata sulla proiezione delle grandezze di sopravvivenza fino al 2040, utilizzando quattro modelli: ARIMA, Lee-Carter, CBD(M6) e APC (Age-Period-Cohort). Questi modelli sono stati scelti per la loro capacità di catturare le tendenze passate e prevedere le dinamiche future, ognuno con caratteristiche uniche che consentono di esplorare diverse possibili evoluzioni della sopravvivenza.

L'implementazione è stata realizzata utilizzando il software statistico R, in particolare attraverso il pacchetto StMoMo (Stochastic Mortality Modeling), che ha permesso di adattare e calibrare in maniera efficiente i modelli sui dati INPS. Questo pacchetto ha migliorato l'accuratezza delle stime e delle proiezioni, offrendo un quadro chiaro delle dinamiche della sopravvivenza.

L'obiettivo finale è quello di valutare l'impatto delle diverse strutture dei modelli sulle proiezioni di sopravvivenza, offrendo indicazioni fondamentali per la gestione delle politiche previdenziali future. Infatti, le previsioni fino al 2040 forniscono uno strumento prezioso per supportare le istituzioni previdenziali e le imprese di assicurazione, nella gestione del rischio di longevità e nella definizione di strategie sostenibili per il futuro del sistema pensionistico.

1 INTRODUZIONE ALLA MORTALITA'.

1.1 Definizione e storia del pensiero sulla morte.

La morte viene definita come la cessazione delle funzioni vitali nell'uomo, negli animali e in ogni altro organismo vivente.

Nel corso della storia umana il tema della morte è stato ampiamente approfondito sia da un punto di vista filosofico che religioso. Nell'antica Grecia, filosofi come Socrate e Platone affrontavano la morte con serenità, vedendola come il momento in cui l'anima si liberava dal corpo e intraprendeva un viaggio verso l'aldilà. Per altri filosofi come Seneca e Epitteto, la morte veniva accettata come parte naturale dell'esistenza umana. Successivamente, con l'avvento del cristianesimo, la morte assunse un significato nuovo, in quanto la fede nella resurrezione e nella vita eterna trasformava il suo significato da fine definitivo a passaggio verso l'aldilà. Questa visione si è mantenuta nel corso dei secoli all'interno della teologia cristiana, fino ad arrivare ai giorni nostri. Nel contesto della filosofia moderna, pensatori come Friedrich Nietzsche hanno messo in discussione le concezioni tradizionali della morte proposte dalla Chiesa Cristiana, ed in particolare nella sua opera "Così parlò Zarathustra" egli propone una visione dell'esistenza umana priva di significato trascendentale, in cui la morte rappresenta la fine assoluta dell'individuo. Infine, altri movimenti religiosi, come ad esempio il Buddhismo, considera la morte come parte di un ciclo continuo di rinascita, noto come "samsara". Secondo questa dottrina, gli esseri senzienti sono destinati a rinascere continuamente in cicli di vita, morte e rinascita, a meno che non raggiungano la liberazione dal ciclo stesso, cioè il "Nirvana". Questi approcci, seppur diversi, offrono un panorama ricco e complesso delle varie interpretazioni umane della morte nel corso della storia.

1.2 La mortalità e i fattori influenti.

Dalla morte deriva il concetto di "mortalità" ed essa, in demografia, rappresenta la principale componente negativa del movimento della popolazione.

In altre parole, la mortalità può essere definita come l'analisi degli aspetti quantitativi che si riferiscono alla morte degli uomini, sia prima della nascita, cioè la mortalità fetale, sia durante il corso della vita, ed essa costituisce un capitolo centrale nella demografia.

Dunque, si osserva che in via generale, la mortalità può essere esaminata come numero assoluto di morti, riferito alla popolazione complessiva (mortalità generica o generale), oppure in senso specifico rispetto a segmenti della popolazione, come la mortalità per sesso, per stato civile, per titolo di studio, per tipo di attività svolta, e così di seguito.

Inoltre, sono molti i fattori che possono influenzare la mortalità. Tra questi è possibile individuare 3 macrocategorie:

- 1) Fattori Socio-Demografici:

- **Età:** l'età è uno dei principali elementi che influenzano la mortalità. Le persone anziane tendono ad avere un rischio più elevato di mortalità rispetto ai giovani. Questo è dovuto al deterioramento delle funzioni fisiche e alla maggiore incidenza di malattie croniche legate all'invecchiamento.
- **Sesso:** esistono differenze di genere nella mortalità. Ad esempio, nelle società occidentali, le donne tendono ad avere un'aspettativa di vita più lunga rispetto agli uomini, tuttavia, gli uomini spesso hanno un rischio più elevato di mortalità per alcune cause, come incidenti stradali o malattie cardiovascolari. Mentre le donne possono avere un rischio più elevato per altre cause, come malattie autoimmuni o osteoporosi.
- **Stato socioeconomico:** il livello di istruzione, il reddito e l'occupazione sono tutti elementi che influiscono sulla mortalità. Infatti, le persone con un basso status socioeconomico possono avere un accesso limitato sia alle risorse sanitarie sia alle cure mediche, il che le porta a vivere in condizioni di vita più precarie affrontando maggiori livelli di stress. Tutti questi sono fattori che aumentano il rischio di mortalità.

2) Fattori Ambientali:

- **Condizioni di vita:** le condizioni abitative, come sovraffollamento, scarsa qualità dell'aria e dell'acqua, mancanza di servizi igienici adeguati e insicurezza alimentare, possono aumentare il rischio di mortalità per malattie infettive, malattie croniche e incidenti.
- **Accesso alle cure mediche:** l'accesso a cure mediche di qualità è fondamentale per ridurre la mortalità. La mancanza di accesso alle cure mediche preventive, agli screening e ai trattamenti può aumentare il rischio di mortalità per malattie prevenibili e trattabili.
- **Stile di vita:** Abitudini come il fumo, l'eccessivo consumo di alcol, una dieta poco salutare, la mancanza di attività fisica e la mancanza di utilizzo di dispositivi di sicurezza, come le cinture di sicurezza in auto, possono aumentare il rischio di mortalità.

3) Fattori Biologici:

- **Malattie:** le malattie croniche, come malattie cardiache, cancro, diabete e malattie respiratorie croniche, sono tra le principali cause di mortalità nel mondo. L'identificazione preventiva dei fattori di rischio e il controllo delle malattie possono contribuire a ridurre il rischio di mortalità.
- **Genetica:** alcune condizioni mediche possono essere influenzate dalla predisposizione genetica. La comprensione del ruolo della genetica nella propensione alle malattie può contribuire a migliorare le strategie di prevenzione e trattamento.

1.3 Storia dello studio mortalità nel mondo.

La prima volta che furono effettuate delle statistiche riguardanti le morti fu nella metà del '600, quando la città di Londra invitò John Graunt a compilare una lista delle statistiche

settimanali delle morti e delle loro cause, che permise di effettuare un'analisi approfondita della mortalità in quella città (v. Graunt,1662).

A questo fine egli costruì la prima “tavola di mortalità” della storia e tracciò il cammino per lo sviluppo delle tecniche di analisi e delle misure della mortalità, fornendo nello stesso tempo le basi per collocare la morte di una persona all'interno del contesto collettivo in cui si verifica.

Gli esperti percepirono immediatamente non solo le potenzialità, per l'igiene pubblica e per lo sviluppo della medicina, dell'informazione statistica sui decessi e sulle loro cause, ma anche l'importanza delle indicazioni che potevano essere desunte circa i legami tra mortalità, condizioni di salute e situazione sociale, ambientale ed economica. Purtroppo, all'epoca non esistevano ancora mezzi e strumenti tali da permettere l'elaborazione di politiche sanitarie in grado di incidere sulle situazioni di rischio e quindi, solo nel Settecento nei paesi più avanzati del Nord Europa, la riduzione dei rischi di morte legati ad alcune malattie iniziò a tradursi in un declino della mortalità complessiva.

Ancora oggi, però, il dibattito sulla genesi della “riduzione della mortalità” rimane molto vivace e le interpretazioni sulle cause che hanno prodotto il suo avvio continuano ad essere discordanti. È certo solo che le prime importanti trasformazioni sono avvenute all'interno del Vecchio Continente e che proprio i paesi europei offrono un laboratorio naturale di analisi dal quale si possono trarre elementi capaci di chiarire i meccanismi che hanno portato agli attuali bassi livelli di mortalità delle popolazioni dei paesi sviluppati.

Il declino della mortalità in Europa si è sviluppato in varie fasi:

1. L'avvio della transizione sanitaria: alle soglie del 700 l'uomo era incapace di contrastare le forze della natura, e quindi la durata media della vita era fortemente condizionata dalle crisi dei mezzi di sussistenza che, provocando periodicamente carestie e profondi stati di miseria, producevano livelli di mortalità elevatissimi. Questa situazione precaria perdurò fino alla metà dell'Ottocento, quando nei paesi dell'Europa centrale e settentrionale, alla perdita di importanza delle crisi si accompagnò un declino sistematico della mortalità; nell'Europa mediterranea e orientale, invece, i livelli erano ancora molto elevati e non sembravano esserci segnali di adeguamento al processo avviato nei paesi all'avanguardia. Così, alle soglie del 900, il vecchio continente presentava tratti tipici di un territorio in cui convivevano situazioni di progresso e di arretratezza, a cui corrispondevano importanti differenze di mortalità (G. Caselli, 1993). In Italia, dal 1880 la mortalità iniziò a ridursi stabilmente, fino a meno del 20% annuo nel 1914. Tale declino, però, fu bruscamente interrotto dall'inizio della Prima guerra mondiale e dalla terribile epidemia di spagnola che provocò un numero di morti altrettanto elevato (G. Caselli, 1990). La differenza tra i paesi del Nord e del Sud Europa era dovuta al ritardo sperimentato da alcuni paesi, e tra questi anche l'Italia, nel dare avvio alla transizione sanitaria. In particolare, il valore della speranza di vita nel periodo antecedente alla transizione sanitaria era di circa 25- 30 anni; successivamente a tale transazione il valore della speranza di vita è arrivato a circa 50- 60 anni.

2. L'aumento della sopravvivenza nel XX secolo: solo a partire dai primi decenni del XX secolo è iniziato anche nei paesi europei meno progrediti un forte declino della mortalità. In quegli stessi anni anche per i paesi più avanzati inizia una nuova fase di rapido declino della mortalità e l'interruzione dell'evoluzione favorevole, provocata dalle crisi di mortalità dovuta alla guerra e all'epidemia di spagnola, viene superata ovunque in tempi brevi. Il passaggio dai vecchi ai nuovi regimi di mortalità si è attuato con tempi e modalità diverse da paese a paese, ma ovunque ha portato le popolazioni europee a sperimentare livelli di sopravvivenza media che sembravano irraggiungibili.

Ciò che si osservava era che, l'aumento della sopravvivenza maschile appariva ovunque più contenuto di quella femminile.

Se le ragioni del rapido declino della mortalità nella prima metà del secolo sono rintracciabili nella sconfitta delle malattie infettive, che aveva permesso la forte riduzione della mortalità nell'età infantile, l'arresto di questa tendenza, e in alcuni casi l'aumento della mortalità negli anni 60, suscitò molte ipotesi interpretative: da un lato si pensò di essere ormai di fronte all'inevitabile rallentamento imposto dalla prossimità ai limiti biologici della specie, dall'altro si iniziò a temere che le crescenti difficoltà fossero imputabili al progressivo deterioramento delle condizioni ambientali al quale gli uomini erano esposti.

3. Miglioramento delle condizioni nel XXI secolo: A livello europeo, nel corso del XXI secolo si è registrato un significativo miglioramento delle condizioni di salute e dell'aspettativa di vita. Secondo l'Organizzazione Mondiale della Sanità (OMS), l'aspettativa di vita in Europa è aumentata di circa 4 anni tra il 2000 e il 2019, passando da circa 75 a circa 79 anni. Questo progresso è stato principalmente dovuto all'accesso migliorato alle cure mediche, all'igiene, all'acqua potabile sicura e ad una migliore nutrizione. Tuttavia, vi sono state differenze significative tra i paesi europei. Ad esempio, paesi come la Norvegia e la Svizzera hanno registrato un aumento maggiore dell'aspettativa di vita rispetto ad altri paesi dell'Europa orientale e meridionale. Anche in Italia, l'aspettativa di vita è aumentata costantemente nel corso del XXI secolo. Secondo l'Istituto Nazionale di Statistica (ISTAT), nel 2000 l'aspettativa di vita media era di circa 78 anni per gli uomini e di circa 83 anni per le donne. Nel 2019, questi numeri sono saliti rispettivamente a circa 81 anni per gli uomini e a circa 85 anni per le donne. Questo aumento è stato attribuito principalmente ai progressi nella sanità pubblica, nell'assistenza sanitaria e nelle tecnologie mediche, che hanno contribuito a ridurre la mortalità da malattie infettive e da altre cause evitabili.

1.4 Elementi di studio della mortalità.

Nel campo delle scienze attuariali, l'analisi della mortalità costituisce un aspetto fondamentale per la valutazione dei rischi e la progettazione dei prodotti assicurativi e piani pensionistici. Per effettuare tale analisi è necessario ricorrere al calcolo di alcuni importanti indicatori statistici, quali:

✚ Tassi di Mortalità:

I tassi di mortalità rappresentano uno strumento cruciale per misurare la frequenza dei decessi all'interno di una popolazione durante un determinato periodo di tempo. Secondo l'Istituto Nazionale di Statistica (ISTAT), il tasso di mortalità è definito come “il rapporto tra il numero di persone decedute in un determinato periodo di tempo e la popolazione media residente in quel periodo, espresso per mille abitanti”.

Tra i principali tassi utilizzati, si parte dall'analisi del “Tasso di Mortalità Grezzo o Crude Death Rate (CDR)”. Questo tasso esprime il numero totale di decessi per mille persone in una popolazione durante un anno, fornendo una visione generale della mortalità. La formula utilizzata per calcolarlo è:

$$CDR = \left(\frac{NUMERO\ TOTALE\ DI\ DECESSI}{POPOLAZIONE\ TOTALE} \right) \times 1000$$

Successivamente, a seconda dell'obiettivo dell'analisi, vengono utilizzati diversi altri tipi di tassi di mortalità, tra cui ad esempio:

- Tasso di Mortalità Specifico per Età: Tale tasso misura il numero di decessi all'interno di specifici gruppi di età ed è utile per valutare il rischio di mortalità associato a diverse fasce d'età. La formula utilizzata per calcolarlo è:

$$Tasso\ di\ Mortalità\ Specifico\ per\ Età = \left(\frac{\text{numero di decessi in un gruppo di età}}{\text{popolazione esposta al rischio in quel gruppo di età}} \right) \times 1000$$

- Tasso di Mortalità Infantile: Esso esprime il numero di decessi di bambini sotto l'anno di età per 1000 nati vivi in un anno, ed è cruciale per valutare la salute neonatale e la qualità delle cure prenatali e post-natali.
- Tasso di Mortalità per Sesso: questo tasso misura il numero di decessi per sesso all'interno di una popolazione, fornendo informazioni sulle differenze di mortalità tra maschi e femmine.
- Tasso di Mortalità Standardizzato per Causa: questo tasso fornisce una stima della mortalità per cause specifiche, come malattie cardiache, tumori, incidenti, ecc. È utile per comprendere le principali cause di morte e per guidare l'allocazione delle risorse sanitarie.
- Tasso di Mortalità degli Anziani: esprime il numero di decessi di persone anziane (solitamente sopra una certa età, ad esempio 65 anni) rispetto alla popolazione anziana esposta al rischio, fornendo indicazioni cruciali sulla salute e sulla longevità della popolazione anziana. Tale tasso è importante perché fornisce alle istituzioni importanti informazioni sugli anziani pensionati. Queste informazioni poi vengono utilizzate per prendere decisioni in ambito previdenziale.

- Tasso di Mortalità per Categoria Occupazionale: questo tasso valuta il rischio di mortalità in base alla categoria occupazionale dell'individuo. Può essere utile per le compagnie assicurative o i fondi pensione nel valutare il rischio di mortalità associato a diverse professioni o settori industriali.

Questo ci fa capire che a seconda dell'analisi che si vuole portare avanti è possibile calcolare diverse tipologie di tassi di mortalità.

✚ Aspettativa di Vita:

secondo l'Organizzazione Mondiale della Sanità (OMS), l'aspettativa di vita è definita come "il numero medio di anni che una persona può aspettarsi di vivere se la struttura per età e sesso della popolazione rimane la stessa nel futuro". L'aspettativa di vita può variare in maniera significativa tra paesi, gruppi demografici e nel corso del tempo, in base a fattori come l'accesso alle cure mediche, lo stile di vita, l'ambiente sociale ed economico, e così via.

Ciò che si osserva è che l'aspettativa di vita, nella maggior parte dei Paesi, è cresciuta molto negli ultimi anni. Tale crescita è dovuta, da un lato, da un miglioramento di alcuni fattori socioeconomici come l'educazione, l'occupazione e la nutrizione, dall'altro dallo sviluppo della medicina e del sistema sanitario, e dallo sviluppo tecnologico che ha permesso alla medicina di raggiungere risultati che fino a qualche anno fa sarebbero sembrati impossibili.

Un ulteriore fattore che ha portato ad un aumento dell'aspettativa di vita è stato un miglioramento della qualità dell'educazione impartita agli individui che, grazie ad essa, hanno acquisito una maggiore consapevolezza nei confronti di uno stile di vita più sano.

1.5 La Tavola di Mortalità di una Generazione.

Quando si affronta lo studio della mortalità, si osserva che l'evento morte che lo caratterizza è un evento definito non rinnovabile, fatale e necessario e, in quanto tale, gestibile con "tavole di eliminazione". Tali tavole rappresentano uno strumento utilizzato per lo studio di fenomeni non rinnovabili (morte, nascita primo figlio, primo matrimonio, ecc) che descrive, secondo l'età, la graduale uscita dallo stato iniziale degli appartenenti ad una generazione a causa dell'evento studiato. Cioè, con una tavola di eliminazione è possibile descrivere come (con quale velocità) la generazione (G) si elimina, totalmente o parzialmente, per effetto dell'evento (E).

➤ TAVOLA DI ELIMINAZIONE:

Età	Ammontare di generazione	Rischio di vivere l'evento	Eventi
X	G_X	r_X	e_X
α	G_α	r_α	e_α		
.	.	.	.		
.	.	.	.		
χ	G_χ	r_χ	e_χ		
.	.	.	.		
.	.	.	.		
.	.	$r_{\beta-1}$	$e_{\beta-1}$		
β	G_β	/	0		

Dove:

- $X - G_X - r_X - e_X$: sono definite "Costanti Biometriche" della tavola
- La colonna (2) nominata "ammontare di generazione" indica i "sopravvivententi", cioè coloro che all'età x non hanno subito l'evento.
- La colonna (3) nominata "rischio di vivere l'evento" indica le probabilità che si verifichi l'evento in ogni classe di età.
- La colonna (4) nominata "eventi" indica l'intensità del fenomeno.
- β = prima età alla quale non si manifesta più l'evento
- G_β = ammontare degli individui che sono definitivamente "sopravvissuti" all'evento. Dato che l'evento "morte" è necessario, in questo caso $G_\beta = 0$.

La "tavola di mortalità" è una particolare "tavola di eliminazione" che descrive, secondo l'età, l'eliminazione per morte (vera o fittizia) di nati, fino all'estinzione dell'ultimo dei componenti. Si avrà, quindi, una progressiva riduzione dell'ammontare iniziale degli individui. Tale strumento è molto utile in quanto misura la cadenza della mortalità al netto di ogni altro fenomeno perturbatore. Cioè, si osserva come si comporta la mortalità alle varie età di una specifica generazione.

Quello che si fa normalmente è porre la radice della tavola (l_0) pari ad una quantità standard, solitamente $l_0 = 100000$, invece che pari all'ammontare reale della

popolazione. Questo ci consente di confrontare le tavole di mortalità di due popolazioni diverse, perché partono entrambe da $l_0 = 100000$.

TAVOLA DI MORTALITA'

Età	Sopravviventi	Probabilità di morte.	Eventi (decessi)
x	l_x	q_x	d_x
0	l_0	q_0	d_0	
1	l_1	q_1	d_1	...
2	l_2	q_2	d_2	...
.
.
x	l_x	q_x	d_x	...
.
.
.
.
.
$\omega-1$	$l_{\omega-1}$	$q_{\omega-1}$	$d_{\omega-1}$...
ω	$l_{\omega} = 0$	/	/	...
				...

Nel quale si osserva che:

- l_x = è la *funzione di sopravvivenza*, cioè il numero di individui sottoposti a rischio di morte tra due compleanni successivi, secondo i valori assunti da q_x .
- q_x = è la *probabilità di morte*, ovvero la probabilità per un individuo di età x , di morire tra x e $x+1$. Esso rappresenta l'elemento cardine della tavola ed è l'unica quantità che è calcolata in base ai dati osservati (dati reali). Per costruire la tavola le q_x devono essere note.
- d_x = è il *numero di decessi* in età tra x e $x+1$.
- ω = è l'età in cui vi è l'estinzione, poiché essendo la morte un evento necessario, non vi è la quota di chi "sfugge" all'evento.

Le principali formule utili per la costruzione della tavola sono:

- $l_0 = 100000$
- $d_0 = l_0 \times q_0 \rightarrow$ (decessi nella prima classe di età)
- $l_1 = l_0 - d_0$
- È una procedura iterativa
- $d_x = l_x \times q_x$
- $l_{x+1} = l_x - d_x$
- $\sum d_x = l_0 = 100000$
- $l_{\omega-1} = d_{\omega-1}$

Un'ulteriore cosa importante da dire è che, i sopravvivenuti, le probabilità e i decessi rappresentano le 3 grandezze fondamentali della tavola e sono legati tra loro da relazioni algebriche per cui basta conoscerne una e si ricavano anche le altre due. La relazione che lega queste 3 grandezze è:

$$q_x = \frac{d_x}{l_x}$$

Ed inoltre, si dimostra che:

- I *decessi* sono "additivi" $\rightarrow d_{x,x+a} = \sum_x^{x+a-1} d_{x,x+1}$
- I *sopravvivenuti* sono "moltiplicativi" $\rightarrow l_{x+a} = \prod_x^{x+a} l_x$
- Le *probabilità* si "combinano".

Successivamente, se si considerasse l'età come una variabile aleatoria continua, si potrebbe osservare come si comportano i *decessi*, i *sopravvivenuti* e le *probabilità di morte* al variare dell'età. Si vanno ad analizzare 3 "curve" molto importanti:

1. CURVA DEI DECESSI:

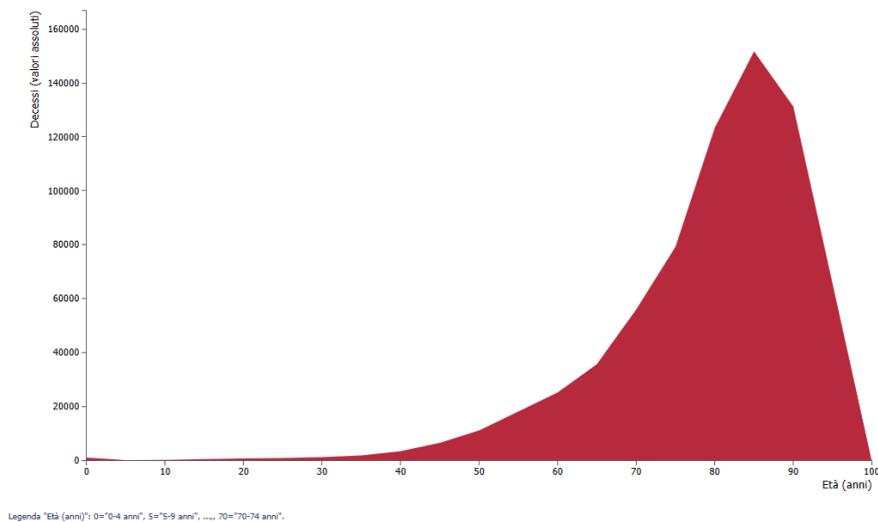


Figura 1: mortalità generale, curva dei decessi. Fonte: MEV(i) - Curva dei decessi (mortalitaevitabile.it)

Con la curva dei decessi viene mostrato il numero di decessi avvenuti per ciascuna età di una determinata generazione di individui. La sua forma ci mostra come il numero di decessi alle età prossime alla nascita sono più elevati rispetto a quelli avvenuti intorno al decimo compleanno, quindi, inizialmente si ha una decrescita fino a circa i 10 anni. Dai 10 anni in poi si osserva una crescita moderata fino ai 40 anni. A partire dai 40 anni tale crescita diventa sempre più forte fino a culminare a circa gli 85 anni di vita. Questo sta a significare che la maggior parte dei decessi viene registrata tra gli 80 e 85 anni. Dagli 85 anni in poi comincia la decrescita che porta alla completa estinzione della generazione.

In particolare, la curva dei decessi è cambiata dal passato al presente e tale cambiamento è dovuto al miglioramento delle condizioni di vita che c'è stato negli anni.

Ciò che si mostra nella curva del presente rispetto a quella del passato è che:

- Si abbassa moltissimo il massimo iniziale in quanto diminuiscono le morti al parto e quelle per mal-nutrizione dei primi mesi di vita.
- Si abbassa il numero di morti in età giovanili e adulte.
- Si sposta in avanti l'età nel quale si concentra il maggior numero di morti, che abbiamo visto essere pari a circa gli 85 anni. Tale età ha un massimo maggiore e quindi c'è una minor dispersione.

2. CURVA della PROBABILITA' DI MORTE:



Figura 2: probabilità di morte per sesso ed età. Italia, anni 1976 e 2016, valori per mille. Fonte: ISTAT.

In linea di massima c'è un comportamento standard che può cambiare di livello a seconda del regime generale di mortalità (se parliamo di una popolazione di un Paese a sviluppo avanzato o no, se si tratta di una generazione del passato o del presente, ecc.)

Ciò che si osserva è che:

- All'età 0 c'è un massimo iniziale dovuto alla mortalità infantile, seguito da una decrescita.
- Ci sono dei piccolissimi massimi in età giovanile che sono dovuti a diversi motivi. In passato questi picchi erano caratteristici delle generazioni femminili perché riguardavano il rischio di morte per parto. Nelle generazioni più recenti, tali picchi sono più caratteristici del sesso maschile e riguardano il rialzo del rischio di morte per incidenti-
- C'è una crescita del rischio di morte in età matura/anziana.
- Con il passare degli anni, il miglioramento delle condizioni di vita ha fatto abbassare la probabilità di morte per tutte le età e quindi questo ha portato ad un conseguente aumento della speranza di vita degli individui.

3. CURVA dei SOPRAVVIVENTI:

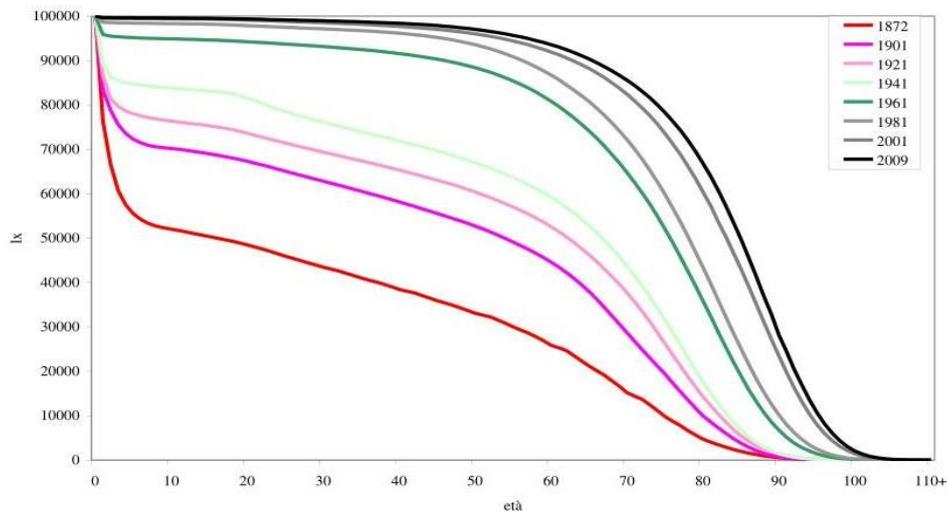


Figura 3: curva dei sopravvivenenti, 1872-2009. Fonte: <http://www.mortality.org/>

La curva di sopravvivenza è una curva decrescente che si articola in tre fasi:

- Il gradino iniziale corrisponde all'elevata probabilità di morte all'età 0.
- Decrescita del numero dei sopravvivenenti.
- Brusca riduzione fino ad arrivare allo 0, ovvero fino ad arrivare alla completa estinzione della generazione.

Più la curva dei sopravvivenenti tende al rettangolo esterno, più le condizioni sono migliori perché vuol dire che l'eliminazione è meno veloce. Nel grafico, la curva peggiore è quella del 1872 nel quale un peso straordinario è rappresentato dalla mortalità infantile e giovanile. La migliore, invece, è quella del 2009 che è caratterizzata da una lentissima eliminazione nei primi anni. Essa inizia a decrescere verso i 60 anni, fino ad arrivare a decrescere molto rapidamente dopo gli 80 anni. Quindi, ciò che è cambiato dal passato al presente è che c'è stato un movimento verso il rettangolo ideale:

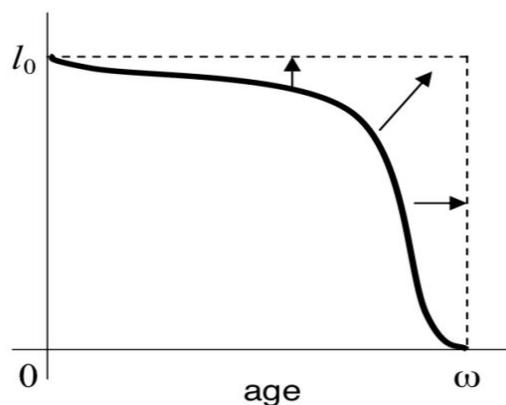


Figura 4: Trend della funzione di sopravvivenza.

Successivamente si osserva come nella tavola di mortalità sono presenti altre “costanti biometriche” oltre a quelle già viste in precedenza, ad esempio:

Età	Sopravvivenuti	P. di morte	Decessi	Rischio di sopravvivenza	Persone - anno	Retrocumulata anni vissuti	Speranza di vita	Vita probabile
X	l_x	q_x	d_x	P_x	L_x	T_x	e_x	π_x
0	100000	q_0	d_0	P_0	L_0	T_0	e_0	π_0
1	l_1	q_1	d_1	P_1	L_1	T_1	e_1	π_1
2	l_2	q_2	d_2	P_2	L_2	T_2	e_2	π_2
.
.
X	l_x	q_x	d_x	P_x	L_x	T_x	e_x	π_x
.
.
$\omega-1$	$l_{\omega-1}$	$q_{\omega-1}$	$d_{\omega-1}$	$P_{\omega-1}$	$L_{\omega-1}$	$T_{\omega-1}$	$e_{\omega-1}$	$\pi_{\omega-1}$
ω	0	/	/	/	/	/	/	/

Dove:

$p_x = 1 - q_x$	Indica la Probabilità di sopravvivenza .
$L_x = \frac{l_x - l_{x+1}}{2}$	Indica il numero di anni vissuti tra l'età x e $x+1$, dagli individui che hanno raggiunto l'età x (persone - anno). In altre parole, essa è una misura del numero di anni vissuto da una generazione all'interno dell'intervallo di età compiuta. La stima di L_x riproduce la popolazione al tempo x : $L_x \approx P_x$
$T_x = \sum_{y=x}^{\omega-1} L_y$	La retrocumulata degli anni vissuti rappresenta la somma degli anni vissuti da tutti gli individui di una popolazione fino all'età x . In altre parole, è la quantità totale di tempo trascorso da tutti gli individui fino ad una certa età.
$e_x = \frac{T_x}{l_x}$	È la speranza di vita , ovvero il n° medio di anni che restano ancora da vivere a ciascuno dei sopravvivenuti all'età x . È un indice sintetico che dà un'idea delle condizioni generali di sopravvivenza della generazione. È un indicatore de cadenza della mortalità, cioè, rappresenta la media ponderata delle età con i decessi.
π_x	Ci si riferisce all' età mediana alla morte .

La speranza di vita emerge come un indicatore di rilievo, particolarmente significativo per le implicazioni legislative nel contesto delle politiche pensionistiche e previdenziali. Tuttavia, per comprendere appieno le dinamiche coinvolte, è cruciale esaminare attentamente la situazione italiana in relazione a questo indicatore.

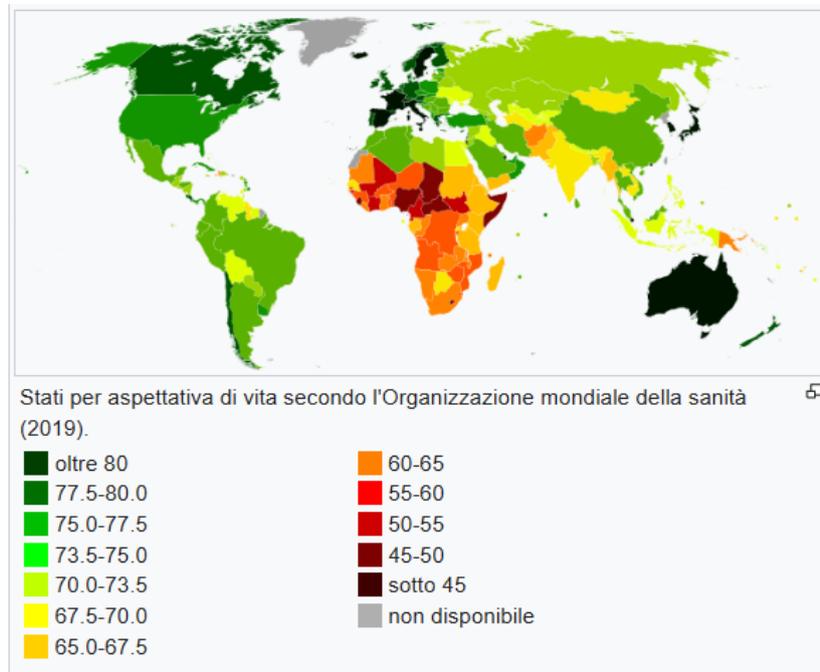


Figura 5: Stati per aspettativa di vita. Fonte: OMS (Organizzazione Mondiale della Sanità).

L'Organizzazione Mondiale della Sanità riconosce l'Italia tra i paesi con un'aspettativa di vita particolarmente alta, fenomeno condiviso con altre nazioni europee come Francia, Spagna e Svezia, che beneficiano di sistemi sanitari avanzati e un tenore di vita elevato. Fuori dall'Europa, l'Australia e soprattutto il Giappone emergono con aspettative di vita notevolmente elevate, il Giappone in particolare vanta il più alto numero di ultracentenari al mondo, fenomeno probabilmente attribuibile a una combinazione di dieta sana, stile di vita attivo e un sistema sanitario efficiente.

D'altro canto, i paesi con aspettative di vita inferiori spesso affrontano sfide legate a condizioni di vita precarie e accesso limitato alle cure mediche. Questa situazione è evidente in molte nazioni dell'Africa centrale e del sud, così come in alcune parti dell'Asia, come India, Pakistan e Afghanistan.

A livello globale, l'aspettativa di vita media si aggira intorno ai 73,3 anni, tuttavia, vi sono notevoli differenze tra i sessi, con gli uomini che registrano una media di 70,8 anni e le donne di 75,9 anni. Secondo i dati del 2015 dell'OMS, le donne tendono a vivere più a lungo degli uomini in quasi tutti i paesi del mondo, eccezion fatta per Mali ed Swatini.

Come ampiamente documentato, nel corso dell'ultimo secolo, vi è stato un notevole aumento della speranza di vita della popolazione. Questo incremento ha provocato un significativo processo di invecchiamento demografico in molte parti del mondo. Grazie ai

progressi sostanziali nella medicina, nelle condizioni di vita e nell'accesso alle cure sanitarie, le persone vivono più a lungo, portando a un aumento proporzionale della popolazione anziana rispetto al totale. Questo fenomeno, noto come invecchiamento della popolazione, presenta nuove sfide e opportunità per i sistemi previdenziali, pensionistici, sanitari ed economici, richiedendo l'adozione di strategie e politiche innovative per garantire il benessere e la sicurezza finanziaria delle attuali e future generazioni.

Tra le varie sfide derivanti dall'invecchiamento della popolazione, emerge in modo significativo il concetto di 'longevity risk', anche definito come rischio di longevità.

2 IL LONGEVITY RISK.

Come già ampiamente detto, la maggior parte del mondo occidentale ha visto, nel corso degli ultimi anni, un costante aumento della vita media dei suoi abitanti. In particolare, Oeppen e Vaupel (2002) ad esempio hanno stimato un aumento di circa tre mesi l'anno dell'aspettativa di vita femminile alla nascita.

Iginio Visco, ex Governatore della Banca d'Italia, in occasione del conferimento della Laurea magistrale ad honorem in Scienze statistiche e attuariali, ha dichiarato che: "vivere più a lungo e in buona salute è un risultato meraviglioso, soprattutto se si hanno le risorse per poterne beneficiare". Allo stesso tempo, sempre Visco I., ha evidenziato come la crescita della longevità pone, tuttavia, anche l'attenzione su una serie di sfide, tra cui il rischio di avere un numero di anziani non autosufficienti. Ciò che rende particolarmente insidioso l'aumento della speranza di vita (cioè, del numero di anni che mediamente ci si può attendere di vivere ancora) è che tale aumento ha sistematicamente superato le previsioni. Tale incertezza sulla lunghezza della vita dà quindi origine a un vero e proprio "rischio di longevità" (longevity risk): esso è un significativo fattore di rischio per i sistemi di previdenza (sia pubblici, sia complementari, sia individuali), che può essere definito ad un livello individuale o aggregato (Stallard 2006):

- A livello individuale, il *longevity risk* si configura come il rischio che l'esistenza degli individui che aderiscono ai sistemi previdenziali, in media, superi la speranza di vita media ipotizzata.
- A livello aggregato, il *longevity risk* si sostanzia nella eventualità che l'insieme dei percettori di rendita vivano, in media, di più rispetto a quanto assunto durante la fase di pricing e quindi, tale rischio, va fortemente a influenzare le rendite pensionistiche e assicurative, facendo sì che venga esteso il periodo di pagamento delle rendite e quindi che possano incrementare le passività.

Attuari e demografi hanno subito riconosciuto i problemi causati dall'invecchiamento della popolazione e dall'aumento della longevità ed hanno quindi dedicato notevole attenzione allo sviluppo di tecniche statistiche utili a studiare tali problematiche. I miglioramenti nei tassi di mortalità sono causati dall'effetto combinato di due fenomeni, che sono la rettangolarizzazione e l'espansione. In particolare, la rettangolarizzazione (particolarmente marcata a partire da 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 ω . L'espansione, invece, consiste nello spostamento in avanti, verso età estreme, del punto di Lexis. Tali marcate tendenze si traducono in un allungamento dell'aspettativa di vita. Si potrebbe quindi dire che il *longevity risk* si materializza quando le aspettative circa la mortalità non si incontrano nell'esperienza della mortalità (M. Johnson 2005). Ciò porta a discutere delle difficoltà che si incontrano nel prevedere la longevità. Infatti, sempre secondo Visco I., i cambiamenti demografici causano deviazioni del numero di morti rispetto ai valori attesi, dove tali deviazioni possono essere sia accidentali che sistematiche. In particolare, le variazioni accidentali costituiscono un vero e proprio rischio assicurativo, in quanto variazioni casuali attorno ad una probabilità di morte nota e predefinita. Tale rischio può essere ripartito tra gli assicurati, con una protezione maggiore al crescere del numero di assicurati. Invece, le variazioni sistematiche costituiscono un rischio collettivo e riguardano differenze sistematiche tra il numero di morti e il suo valore atteso a livello di coorte.

Le varie fonti di rischio responsabili sia delle variazioni accidentali che di quelle sistematiche 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 temporale che governa l'evoluzione della mortalità;

La natura sistematica del rischio di longevità rende impraticabile l'uso di operazioni di diversificazione del portafoglio. Le tecniche sviluppate finora per affrontare questo rischio si possono essenzialmente ricondurre a due approcci principali: la gestione autonoma del rischio e il trasferimento del rischio a terze parti. In particolare, nel primo caso, le compagnie assicurative utilizzano tavole di mortalità proiettate.

La letteratura statistico-attuariale ha sviluppato vari modelli estrapolativi al fine di proiettare i tassi di mortalità. Tali modelli sono quelli più utilizzati dagli attuari e permettono di individuare schemi ricorrenti e prevedere l'andamento futuro di una certa variabile, partendo da una serie storica di dati relativi a quella variabile, come nel caso dei trend storici della mortalità. Essi si fondano sull'assunzione che l'evoluzione della mortalità osservata in passato si possa ripetere in futuro.

Tra i modelli estrapolativi è possibile effettuare una distinzione tra quelli deterministici e quelli stocastici:

- *Deterministici*: proiettano le tendenze di sopravvivenza passate verso il futuro. In questo caso la natura stocastica della mortalità non viene considerata.

- *Stocastici*: nei quali i tassi di mortalità osservati sono considerati come risultati di variabili aleatorie che rappresentano la mortalità passata, mentre i tassi di mortalità proiettati sono stime di variabili aleatorie che rappresentano la mortalità futura. Quindi si può dire che nel caso dei modelli stocastici le previsioni sono ottenute sulla base di distribuzioni di probabilità. Quindi, in un modello che collega la struttura probabilistica del processo stocastico a un insieme di dati osservati, si passa dai valori osservati della frequenza di mortalità casuale a una possibile proiezione futura di questa.

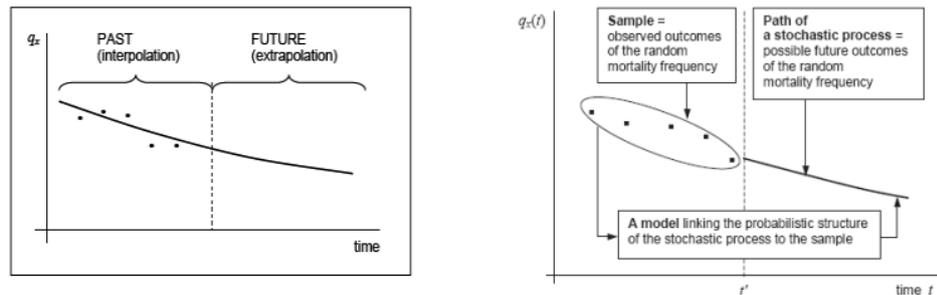


Figura 6: modelli stocastici di mortalità. Fonte: Olivieri-Pitacco “Life annuities and longevity dynamics” WP n. 36, CERAP 2006

Un modello stocastico di questo tipo, dunque, permette di ottenere sia stime precise che intervalli di confidenza per i futuri tassi di mortalità.

Due famiglie di modelli molto utilizzate in ambito attuariale sono:

- Lee-Carter, 1992, più successive estensioni;
- Cairns-Blake-Dowd, 2006, più successive estensioni;

Senza dubbio, il modello più importante per la modellizzazione dei tassi di mortalità è il modello proposto da due noti demografi e statistici statunitensi, ovvero Ronald Lee e Lawrence Carter. Il modello in questione, infatti, prende il nome di *Modello Lee-Carter* (1992). Successivamente, tale modello ha ispirato numerose varianti e estensioni, come ad esempio: *Lee e Miller (2001)*, *Booth et al. (2002)* e *Brouhns et al. (2002)*. L’obiettivo di tali estensioni è quello di migliorare la bontà di adattamento e la previsione del modello.

Inoltre, il modello a due fattori Cairns-Blake-Dowd (CBD), presentato da Cairns et al. nel 2006, rappresenta una delle principali varianti del modello Lee-Carter. In particolare, esso assume che, il logit della probabilità di morte di un anno sia funzione lineare dell’età, considerando i parametri di intercetta e pendenza, attraverso gli anni, dei processi stocastici. Successivamente, anche il Modello CBD ha avuto tre importanti estensioni.

Considerata la crescente varietà e il rapido aumento dei modelli stocastici di mortalità presentati nella letteratura, si è cercato di identificare elementi comuni tra di essi. Hunt e Blake (2015) hanno esaminato le caratteristiche dei modelli di mortalità e hanno delineato una struttura basata sui concetti di età, periodo e coorte, che include la maggior parte dei metodi stocastici di mortalità.

2.1 La Notazione.

Prima di osservare nel dettaglio i vari modelli è importante indicare quelle che sono le notazioni con il quale tali modelli operano.

Verranno indicati con:

- $D_{x,t}$: la variabile aleatoria che rappresenta il n° di decessi in una popolazione all'età x e durante l'anno di calendario t ;
- $d_{x,t}$: è il n° osservato di decessi;
- $E_{x,t}^c$: è l'esposizione centrale al rischio all'età x , nell' anno t ;
- $E_{x,t}^0$: è l'esposizione iniziale al rischio;

Grazie a tali grandezze è possibile determinare la probabilità di morte mono-annuale per un individuo di età x nell'anni di calendario t . Questa grandezza può essere stimata nel modo seguente:

$$\hat{q}_{x,t} = \frac{d_{x,t}}{E_{x,t}^0}$$

Inoltre, altre grandezze importanti sono:

- $\mu_{x,t}$: è la forza di mortalità;
- $m_{x,t}$: è il tasso centrale di mortalità; la sua stima empirica è ottenuta come: $\hat{m}_{x,t} = \frac{d_{x,t}}{E_{x,t}^c}$;

Sotto l'ipotesi che la forza di mortalità $\mu_{x,t}$ sia costante per ogni età x e per ogni anno di calendario t , cioè dall'età x all'età $x+1$, si osserva che la forza di mortalità e il tasso centrale di mortalità coincidono. Quindi $\mu_{x,t} = m_{x,t}$.

Utilizzando, all'interno del software *R*, il pacchetto *StMoMo*, si può assumere che i decessi $d_{x,t}$, le esposizioni centrali al rischio $E_{x,t}^c$ o le esposizioni iniziali al rischio $E_{x,t}^0$, siano disponibili in un formato di allineamento rettangolare comprendente le età (sulle righe) e gli anni di calendario (sulle colonne). Una relazione molto importante che intercorre tra queste grandezze dice che quando sono disponibili solo le esposizioni centrali al rischio e sono richieste quelle iniziali (o viceversa), è possibile approssimare le esposizioni centrali tramite l'aggiunta della metà del numero dei decessi (relativo a quelle che sono le esposizioni centrali), cioè:

$$E_{x,t}^0 \approx E_{x,t}^c + \frac{1}{2} d_{x,t}$$

Nel caso in cui il contesto di analisi risulta essere ben definito, è possibile scrivere $E_{x,t}$ per riferirsi a $E_{x,t}^0$ oppure a $E_{x,t}^c$.

2.2 Modelli stocastici generalizzati di mortalità di tipo APC (Age-Period-Cohort).

Come evidenziato all'interno dell'articolo intitolato "StMoMo: An R Package for Stochastic Mortality Modeling" pubblicato da A.M. Villegas, P. Millosovich e V.K. Kaishev, alcuni autori hanno recentemente cercato di identificare le somiglianze tra i vari modelli di mortalità stocastica.

Ad esempio, Hunt e Blake (2015) descrivono una struttura di tipo "Age-Period-Cohort" (APC) che comprende molti modelli di mortalità stocastica. In un altro contributo rilevante, Currie (2016) dimostra che molti modelli di mortalità comunemente usati possono essere formulati utilizzando la terminologia standard dei modelli lineari o non lineari generalizzati. In questa sezione, è possibile mantenere l'attenzione su quelli che sono i modelli stocastici di mortalità per età-coorte generalizzati (GAPC).

Così come i modelli lineari generalizzati, un modello GAPC è composto da quattro elementi (McCullagh e Nelder, 1989):

1. La componente casuale: il numero di decessi $D_{x,t}$ segue una distribuzione di Poisson o una distribuzione binomiale, in modo che:

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

oppure

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

con

$$E\left(\frac{D_{x,t}}{E_{x,t}^c}\right) = \mu_{x,t}$$

e

$$E\left(\frac{D_{x,t}}{E_{x,t}^0}\right) = q_{x,t}$$

rispettivamente.

2. La componente sistematica: seguendo quanto detto da hunt e Blake nel 2015, gli effetti dell'età x , dell'anno di calendario t e dell'anno di nascita (coorte) $c = t - x$, sono catturati attraverso un predittore $\eta_{x,t}$ dato da:

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

Dove:

- Il termine a_x è una funzione statica dell'età che cattura la forma generale della mortalità per età. In altre parole, mostra 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 $\kappa_t^{(i)}$, con $i = 1, \dots, N$, contribuisce a spiegare la variazione della mortalità nel tempo e $\beta_x^{(i)}$ invece descrive per ogni età come la mortalità reagisce al variare di $\kappa_t^{(i)}$.

- Il termine γ_{t-x} (indice di coorte) tiene conto dell'effetto di coorte con $\beta_x^{(0)}$ che descrive per ogni età come la mortalità varia al variare dell'effetto coorte.

I termini $\beta_x^{(i)}$, con $i = 0, 1, \dots, N$, possono essere sia funzioni predefinite dell'età, come ad esempio $\beta_x^{(i)} \equiv f^i(x)$, come accade nei modelli di tipo CBD, sia termini parametrici senza alcuna struttura che deve essere stimata, come avviene nel modello Lee Carter.

Nella famiglia GAPC è possibile assumere che gli indici di periodo $\kappa_t^{(i)}$, con $i = 1, \dots, N$, e l'indice di coorte γ_{t-x} , siano dei processi stocastici. Essa va a rappresentare la caratteristica principale che consente la proiezione stocastica dei modelli GAPC.

3. La link function (funzione di collegamento) g , è l'elemento che associa la componente casuale a quella sistematica, tale che:

$$g\left(E\left(\frac{D_{x,t}}{E_{x,t}}\right)\right) = \eta_{x,t}$$

È possibile utilizzare diverse funzioni di collegamento, ma è comune optare per il collegamento canonico. In questo caso, la distribuzione di Poisson è tipicamente associata alla funzione di collegamento *logaritmica*, mentre la distribuzione binomiale è comunemente abbinata alla funzione di collegamento *logit*. Per una trattazione dettagliata di questi aspetti nel contesto dei modelli di mortalità, si può fare riferimento a Currie (2016), mentre McCullagh e Nelder (1989) offrono una panoramica più generale nell'ambito dei modelli lineari generalizzati (GLM).

4. L'insieme dei vincoli sui parametri: la maggior parte dei modelli stocastici, che permettono lo studio della mortalità, sono identificabili solo attraverso una serie di vincoli sui parametri che permettono la calibrazione. In assenza di tali vincoli ci si imbatte in una serie di problemi di identificabilità dei parametri. I vincoli sopracitati vengono applicati tramite una "funzione di vincolo" identificata con v , che va ad associare un vettore arbitrario di parametri ad un vettore di parametri trasformati, soddisfacendo i vincoli del modello senza alcun effetto sul predittore $\eta_{x,t}$.

Il vettore arbitrario è il seguente:

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

Il vettore di parametri trasformati è il seguente:

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

In conclusione, si osserva che:

θ e $\tilde{\theta}$ producono lo stesso $\eta_{x,t}$.

La maggior parte dei modelli di mortalità stocastica proposti appartengono alla famiglia dei modelli GAPC. Sono inclusi tra questi il modello originale Lee-Carter più le relative estensioni come, ad esempio, quelle proposte da Renshaw e Haberman (2003,2006), il modello CBD originale e le estensioni proposte da Cairns et al. (2009). Oltre a questi, anche tutte le strutture dei modelli considerate in Haberman e Renshaw (2011), Lovász (2011) e van Berkum et al. (2014), e anche i

modelli proposti da Plat (2009), Aro e Pennanen (2011), O'Hare e Li (2012), Borger et al. (2013) e Alai e Sherris (2014), fanno parte dei modelli GAPC.

2.3 Il Modello Lee-Carter (1992). (M1)

I modelli che vengono utilizzati per lo studio della mortalità, come è stato già detto in precedenza, sono molteplici. Il primo che viene analizzato, che è contraddistinto con la sigla "M1" è il Modello Lee-Carter.

Il modello Lee-Carter è un modello statistico-demografico utilizzato per la previsione dei tassi di mortalità. Tale modello prende il nome da Ronald Lee e Lawrence Carter che lo introdussero nel 1992 tramite un articolo intitolato "Modeling and forecasting the time series of U.S. mortality" pubblicato sul "Journal of the American Statistical Association".

La prima applicazione di tale modello è stata effettuata, proprio dagli autori, utilizzando dati sulla mortalità della popolazione degli Stati Uniti d'America nel periodo che va dal 1933 al 1987. Tuttavia, il metodo viene attualmente utilizzato per tutte le cause e per i dati di mortalità specifici per cause provenienti da diversi paesi e periodi di tempo, andando ben oltre la sua applicazione originaria.

Il modello Lee-Carter ha il vantaggio di essere molto semplice da un punto di vista applicativo, ma allo stesso tempo gode della capacità di descrivere in maniera molto accurata il fenomeno della mortalità.

Altra caratteristica fondamentale di questo modello è che esso è puramente estrapolativo. Ciò sta a significare che esso non va a inglobare informazioni sociali, comportamentali o mediche, ma tutte le considerazioni sono basate su come il fenomeno si è comportato in passato, assumendo che tale comportamento, poi possa ripetersi in futuro. Infatti, il punto di forza di tale modello è proprio la combinazione di un modello demografico con metodi statistici sulle serie storiche.

Inoltre, così come descritto anche durante la "notazione" vista in precedenza, si osserva che anche tale modello appartiene a quella classe di modelli nei quali l'andamento della mortalità viene descritto sia in funzione del tempo t sia in funzione dell'età x .

Il modello, in particolare, ipotizza un andamento esponenziale del tasso centrale di mortalità $m_{x,t}$ per ogni età. Tale tasso viene considerato come una probabilità di decesso calcolata alla metà del periodo in considerazione, attraverso il rapporto tra i decessi (di età x nell'anno t) e il numero medio di sopravvissuti alla metà del periodo considerato. Esso si differenzia da q_x , ovvero dalla probabilità di decesso a inizio periodo, che invece si calcola come il rapporto tra i decessi e il numero di persone a inizio periodo.

Nella versione originale, si osserva che il tasso centrale di mortalità $m_{x,t}$, per l'età x ($x = x_1, \dots, x_n$) e per l'anno di calendario t ($t = t_1, \dots, t_n$), è definito tramite la seguente relazione log-bilineare:

$$(1) \quad \ln m_{x,t} = a_x + \beta_x \kappa_t + \varepsilon_{x,t}$$

Oppure, definito anche come:

$$(2) \quad m_{x,t} = e^{a_x + \beta_x \kappa_t + \varepsilon_{x,t}}$$

Dove:

- a_x : è il parametro che descrive in media il comportamento del tasso centrale di mortalità al variare dell'età;
- β_x : è il parametro che descrive, per ogni età, come la mortalità reagisce al variare di κ_t ;
- κ_t : è l'indice della variazione della mortalità nel tempo. Esso mostra, per tutte le età insieme, come il fenomeno della mortalità si è evoluto nel passato;
- $\varepsilon_{x,t}$: rappresenta il termine di errore, cioè la parte di mortalità che non è catturata dal modello. Tali errori risultano essere I.I.D. (identicamente e indipendentemente distribuiti) e anche normalmente distribuiti: $\varepsilon_{x,t} \sim N(0, \sigma_\varepsilon^2)$.

Si osserva che a_x , β_x e κ_t sono parametri da stimare, mentre $\varepsilon_{x,t}$ è un insieme di termini di disturbo casuali.

La parametrizzazione (1) non è univoca, in quanto rimane invariata sotto determinate trasformazioni, ovvero:

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

$$a_x \rightarrow a_x - c\beta_x \quad \kappa_t \rightarrow \kappa_t + c \quad \forall c \in \mathbb{R}$$

Tuttavia, questo non costituisce un ostacolo, ma significa semplicemente che la probabilità associata al modello ha un numero infinito di massimi equivalenti, ognuno dei quali potrebbe produrre previsioni identiche. In pratica, come dichiarato da F.Girosi e G.King (2007), ciò che si fa è limitare la scelta ad una parametrizzazione arbitraria ma coerente per l'identificazione. Questo può essere fatto imponendo 2 vincoli sui parametri molto importanti, in assenza dei quali ci sarebbero problemi di identificazione dei parametri stessi.

Tali vincoli sono:

$$\sum_x \beta_x = 1$$

$$\sum_t \kappa_t = 0$$

Successivamente si va a semplificare la notazione andando a considerare con $m_{x,t}$, direttamente il logaritmo del tasso centrale di mortalità.

Il secondo vincolo imposto è molto importante in quanto implica immediatamente che il parametro a_x è semplicemente la media empirica di $m_{x,t}$, lungo l'intero periodo di osservazione, ovvero:

$$a_x = \bar{m}_x$$

Riscriviamo quindi il modello in termini di tasso medio di mortalità logaritmica centrato:

$$\bar{m}_{x,t} = m_{x,t} - \bar{m}_x$$

Successivamente, poiché gli usi pratici del modello Lee-Carter assumono implicitamente che i termini di disturbo ($\varepsilon_{x,t}$) siano normalmente distribuiti e allora è possibile riscrivere l'equazione (1) come un modello moltiplicativo a effetti fissi per il profilo di età centrato:

$$(3) \quad \bar{m}_{x,t} \sim N(\bar{\mu}_{x,t}, \sigma^2)$$

$$E(\bar{m}_{x,t}) \equiv \bar{\mu}_{x,t} = \beta_x \kappa_t$$

Tramite questa espressione, verranno utilizzati solo $X+T$ parametri per approssimare gli $A \times T$ elementi della sella seguente matrice:

	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}
⋮	⋮	⋮	⋮	⋮	⋮	⋮
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	

Figura 7: Fonte: Girosi-King "Understanding the Lee-Carter Mortality Forecasting Method", 14 Settembre 2007

Per esempio, Lee-Carter approssima $\bar{m}_{5,0}$ (nella cella in alto a sinistra) con il prodotto dei parametri alla fine della prima riga e della prima colonna $\beta_5 \kappa_0$.

Sotto questo aspetto il modello Lee-Carter può anche essere pensato come un caso speciale di modelli Log-Lineari per tabelle di contingenza (Bishop, Fienberg e Holland, 1975; King, 1989: cap.6), dove i valori delle celle sono approssimati con stime dei parametri che rappresentano i marginali.

In effetti, questo modello è la versione più semplice dei modelli per tabelle di contingenza, in cui si assume l'indipendenza delle righe (gruppi di età) e delle colonne (periodi di

tempo), ed in cui il valore atteso della cella è dato semplicemente dal prodotto dei parametri ricavati dalle rispettive marginali:

$$E(\bar{m}_{x,t}) = \beta_x \kappa_t$$

In un modello di tabella di contingenza, questa ipotesi sarebbe appropriata se la variabile rappresentata sulle righe della matrice fosse indipendente da quella osservata sulle colonne.

Tale ipotesi, per il tasso di log-mortalità, significa che bisogna assumere l'assenza di interazioni (indipendenza) tra l'età ed il tempo, e cioè ipotizzare che i parametri β_x e κ_t rimangono fissi nel tempo per ogni età x e per ogni tempo t , rispettivamente.

Questa però è un'assunzione poco realistica soprattutto quando si considerano età elevate, nel quale si osserva una maggiore variabilità della mortalità. Proprio per tale ragione vengono proposte una serie di estensioni del modello Lee Carter, che consentono, oltre a cambiamenti strutturali, anche variazioni di β_x nel tempo.

2.3.1 Stima dei parametri.

I parametri β_x e κ_t del modello possono essere stimati tramite il cosiddetto "metodo della massima verosimiglianza". Tuttavia, i vincoli che sono stati imposti causano malfunzionamenti dei programmi di ottimizzazione standard. Lee e Carter, però, sottolineano come la soluzione ottimale può essere trovata attraverso la scomposizione in componenti principali, anche definita Singular Value Decomposition (SVD), della matrice centrata dei profili di età, $\tilde{m} = BLU'$, dove la stima per β_x è la prima colonna di B , mentre la stima per κ_t è $\beta' \tilde{m}_t$. Se la decomposizione SVD di \tilde{m} non è disponibile si può calcolare come auto vettore normalizzato della matrice $C \equiv \tilde{m}\tilde{m}'$, corrispondente all'autovalore più grande. Indipendentemente dal fatto che si utilizzi la SVD o gli autovalori per trovare l'ottimo, la giustificazione teorica della procedura rimane il metodo della massima verosimiglianza. Nella pratica però, Lee e Carter suggeriscono che, dopo che β e κ sono stati stimati, si osserva che il parametro κ_t può essere ristimato utilizzando un criterio diverso. Tale fase di rivalutazione, spesso chiamata "second stage estimation", non sempre ha una soluzione unica per il criterio evidenziato da Lee e Carter nel 1992. Inoltre, più recentemente sono stati proposti una serie di criteri diversi (Lee e Miller, 2001; Wilmoth, 1993), e alcuni ricercatori saltano del tutto questa fase di ristima.

Più nel dettaglio, per attuare la scomposizione in valori singolari (SVD), occorre definire una funzione obiettivo da minimizzare tramite il metodo dei minimi quadrati (OLS). Tale funzione è la seguente:

$$\vartheta_{LS}(\mathbf{a}, \boldsymbol{\beta}, \boldsymbol{\kappa}) = \sum_{x=x_1}^{x_n} \sum_{t=t_1}^{t_n} (\ln \hat{m}_x(t) - a_x - \beta_x \kappa_t)^2$$

dove, si osserva che $\hat{m}_x(t)$ indicano i tassi osservati.

Successivamente, si effettua il calcolo della derivata parziale di tale funzione rispetto a a_x . Tale derivata viene posta uguale a 0, e si ottiene che:

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

Siccome, come già detto in precedenza, per il modello Lee Carter vale il vincolo $\sum_t \kappa_t = 0$, allora si ottiene che:

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

Invece, per quanto riguarda i termini β e κ , si va ad effettuare la "scomposizione in componenti principali" del primo termine della matrice centrata $\ln \hat{m}_x(t) - \hat{a}_x$.

Si consideri la matrice dei tassi centrali di mortalità:

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

e poi si definisce la matrice da minimizzare:

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

I parametri β_x e κ_t vengono scelti in maniera tale da minimizzare la seguente funzione:

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

Infine, occorre individuare gli autovettori e gli autovalori della matrice \mathbf{Z} .

Successivamente, per la relazione descritta dall'equazione (1) e per l'ipotesi di normalità degli errori, si osserva che:

$$(4) \quad \ln m_{x,t} - a_x = \beta_x \kappa_t + \varepsilon_{x,t} \sim N(\beta_x \kappa_t, \sigma_x^2)$$

Come già menzionato in precedenza, Lee e Carter scelgono di utilizzare il metodo della decomposizione ai valori singolari, nonostante questo non sia sempre integrato nei pacchetti software statistici. In alternativa, si effettua l'utilizzo del metodo della massima verosimiglianza, con l'obiettivo di stimare il seguente prodotto: $\beta_x \kappa_t$.

Inoltre, siccome gli errori sono normalmente distribuiti e siccome vale il vincolo $\sum_x \beta_x = 1$, si osserva come l'indice κ_t può essere stimato nel seguente modo:

$$\hat{\kappa}_t = \sum_{x=x_1}^{x_n} \ln \hat{m}_{x,t}$$

Poi, considerando la relazione data dall'equazione (4), il parametro β_x può essere stimato tramite un modello di regressione lineare su κ_t (separatamente per ogni gruppo di età x), cioè:

$$\ln \hat{m}_{x,t} = \beta_x \hat{\kappa}_t$$

A questo punto, per ogni valore di κ , il modello va a definire un insieme costituito da una serie di tassi centrali di decesso, che possono essere utilizzati per derivare una tavola di mortalità. Tali tassi sono ricavati dalla seguente relazione:

$$q_{x,t} \approx \frac{2m_{x,t}}{2 + m_{x,t}}$$

oppure:

$$q_{x,t} \approx 1 - e^{-m_{x,t}}$$

La procedura, descritta in questo modo, può essere eseguita esplorando un intervallo di valori accettabili di κ . Non esiste, infatti, una soluzione analitica, poiché le stime di $\hat{\beta}_x$ e di $\hat{\kappa}_t$, non garantiscono che il numero di decessi stimato sia uguale al numero di decessi osservato nel tempo.

Inoltre, l'ipotesi di normalità degli errori, e quindi dei tassi centrali di log-mortalità, potrebbe non essere pienamente supportata dai dati campionari.

Pertanto, gli autori propongono l'utilizzo di una procedura iterativa che permette di far sì che il numero di decessi stimato sia approssimativamente uguale al numero di decessi osservato.

A questo punto, occorre dire che il metodo appena descritto non ammette una soluzione unica, e nel caso in cui ci sono molteplici soluzioni, o nel caso in cui non c'è soluzione, allora a quel punto il modello proposto da Lee e Carter diviene inconsistente.

Proprio per questo è importante dire che il metodo Lee-Carter presenta 3 limiti da non sottovalutare:

- Il modello necessita dei vincoli sui parametri β_x e κ_t per poter essere calibrato. In assenza di tali vincoli si creerebbe un problema di identificabilità dei suddetti parametri.
- Il modello potrebbe risultare essere non efficiente per le età avanzate a causa di una scarsa disponibilità di dati.
- Il parametro β_x , potrebbe risultare essere negativo per alcune età. Ciò indica che per tali età la mortalità tende ad aumentare.

Di contro, il modello proposto da Lee-Carter ha anche altrettanti vantaggi, che sono:

- Il suo ampio utilizzo sia in letteratura che in pratica lo ha fatto diventare uno dei modelli principali, se non il principale, nell'ambito dello studio della mortalità.

- Un altro vantaggio è che tale modello necessita di un numero esiguo di parametri dal significato demografico semplice per essere utilizzato. Ciò comporta una facilità applicativa ed interpretativa.
- Un ulteriore vantaggio è dato dal fatto che, per un dato valore di κ_t , viene definito un set di probabilità di morte che consente la derivazione di un'intera tavola di mortalità.

A questo punto, una volta spiegato cos'è il modello Lee-Carter e come è possibile stimare i parametri, è possibile parlare della fase relativa al "forecasting", ovvero alla proiezione e previsione dei futuri tassi di mortalità.

2.3.2 Forecasting (proiezione dei parametri).

Per quanto riguarda le previsioni della mortalità, viene ipotizzato che le grandezze a_x e β_x rimangono costanti (essi, infatti, possono essere considerati indipendenti dal tempo).

Proprio per questo sarà necessario proiettare solo ed esclusivamente il parametro dipendente dal tempo, ovvero $\hat{\kappa}_t$. Tale parametro viene considerato e modellizzato come un processo stocastico tramite la procedura *Box-Jenkins*. Tale procedura è stata proposta dagli studiosi Box e Jenkins nel 1976, e si basa su 3 passaggi fondamentali, che sono: 1) l'identificazione; 2) la stima dei parametri; 3) il controllo diagnostico. L'utilizzo della procedura Box-Jenkins è importante perché consente l'individuazione e la generazione di un adeguato modello $ARIMA(p,d,q)$ per l'indice di mortalità κ_t .

La sigla ARIMA sta per AutoRegressive Integrated Moving Average, in italiano definito processo AutoRegressivo Integrato a media mobile, e va a rappresentare un'estensione, proposta da Box e Jenkins, dei processi ARMA. Tale estensione è finalizzata a considerare anche processi non stazionari, che però possono essere resi stazionari attraverso opportune differenziazioni.

In generale, i modelli di serie storiche $ARIMA(p,d,q)$ sono descritti dalla seguente equazione:

$$\nabla^d Y_t = \delta + \phi_1 \nabla^d Y_{t-1} + \dots + \phi_p \nabla^d Y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q}$$

Dove:

- Il parametro p indica l'ordine della componente autoregressiva;
- Il parametro d indica il numero di differenziazioni effettuate per rendere il processo stazionario;
- Il parametro q indica l'ordine della componente della media mobile;

Come però è stato evidenziato da Federico Girosi e da Gary King nell'elaborato già sopracitato "Understanding the Lee-Carter Mortality Forecasting Method", dopo aver testato molti modelli ARIMA, è stato scoperto che il modello più adeguato risulta essere un "random walk with drift" (il termine "drift" sta a indicare un termine di "deriva"). In particolare, gli autori ci tengono a sottolineare che, in teoria, altri modelli ARIMA

potrebbero essere preferibili per set di dati diversi, ma in pratica, è stato utilizzato quasi esclusivamente il modello “random walk with drift” per κ_t .

Il modello del quale si parla nell’elaborato è il seguente:

$$\begin{aligned}\hat{\kappa}_t &= \hat{\kappa}_{t-1} + \theta + \xi_t \\ \xi_t &\sim N(0, \sigma_{rw}^2)\end{aligned}$$

dove θ rappresenta il termine di deriva (drift) e la sua stima di massima verosimiglianza è semplicemente:

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

che dipende solo dalla prima e dall’ultima stima di κ .

Invece, il termine ξ_t rappresenta il termine di errore nel tempo, ed esso risulta essere normalmente distribuito con media nulla e varianza pari a σ_{rw}^2 . La stima di tale varianza è la seguente:

$$\hat{\sigma}_{rw}^2 = \frac{1}{T - 1} \sum_{t=1}^{T-1} (\hat{\kappa}_{t+1} - \hat{\kappa}_t - \hat{\theta})^2$$

con:

$$Var[\hat{\theta}] = \frac{\sigma_{rw}^2}{T - 1}$$

Nel caso in cui si volessero prevedere due periodi in anticipo, viene inserita la stima del parametro di deriva $\hat{\theta}$ e si va a sostituire la definizione di $\hat{\kappa}_{t-1}$ spostato indietro nel tempo di un periodo, cioè:

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

dove:

$$\hat{\kappa}_{t-1} = \hat{\kappa}_{t-2} + \hat{\theta} + \xi_{t-1}$$

e quindi, per 2 periodi in anticipo, si avrà che:

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

Alla luce di ciò, la previsione dell’indice di mortalità proiettato al generico istante $T+(\Delta t)$, con $\Delta t = 1, 2, \dots$, e calcolato con i dati disponibili fino all’istante T , si ottiene nel modo seguente:

$$(5) \quad \hat{\kappa}_{T+(\Delta t)} = \hat{\kappa}_T + (\Delta t)\hat{\theta} + \sum_{l=1}^{\Delta t} \xi_{T+l-1}$$

che, in virtù dell'ipotesi di indipendenza e di identica distribuzione degli errori ξ_t , diventa:

$$= \hat{\kappa}_T + (\Delta t)\hat{\theta} + \sqrt{(\Delta t)}\xi_t$$

Successivamente, utilizzando le condizioni dettate dall'equazione (2) e dall'equazione (5), è possibile ricavare le previsioni dei futuri tassi di mortalità per ogni età x e per ogni istante Δt . Tale previsione è data dalla seguente espressione:

$$(6) \quad m_{x,t_n+\Delta t} = e^{\hat{\alpha}_x + \hat{\beta}_x(\hat{\kappa}_{t_n} + \Delta t\hat{\theta} + \sqrt{\Delta t}\xi_{t_n}) + \epsilon_{x,t_n+\Delta t}}$$

Risulta opportuno sottolineare che nell'equazione (6) sono presenti due tipologie di errore che sono: un errore di proiezione ed un errore derivante dal fitting dei dati, nel quale il secondo è dovuto alla natura puramente estrapolativa non solo delle proiezioni ma anche del modello in esame.

Successivamente, si osserva come nel corso del tempo sono state sviluppate molte estensioni e generalizzazioni del modello Lee-Carter. Tali estensioni, seppur introducono nuovi parametri, effetti di periodo o effetti coorte, comunque soffrono delle stesse problematiche. Una delle più importanti estensioni del modello Lee-Carter è quella proposta da Brouhns et al.

2.3.3 Estensione di Brouhns et al. (2002).

Nel 2002, N.Brouhns, M.Denuit e JK. Vermunt, all'interno dell'elaborato intitolato "A Poisson log-bilinear regression approach to the construction of projected lifetables", hanno proposto un'estensione molto importante del modello Lee-Carter.

Secondo Alho (2000), il modello descritto da Lee-Carter non è propriamente adatto, come già accennato, lo svantaggio principale della stima tramite la SVD (Singular Value Decomposition) è che si presume che gli errori siano omoschedastici, cioè che siano caratterizzati dalla stessa varianza al cambiare dell'età x e dell'anno di calendario t . Ciò è legato al fatto che si sta assumendo che gli errori siano normalmente distribuiti, il che risulta essere abbastanza irrealistico. Infatti, il logaritmo della forza di mortalità risulta essere molto più variabile in età avanzate rispetto a quelle giovani, a causa del numero assoluto di decessi molto più piccolo in età avanzata. Tale fenomeno porta al problema chiamato "chiusura della tavola".

Poiché il numero di decessi è una variabile casuale di conteggio, secondo Brillinger (1986), l'ipotesi di Poisson sembra essere la più plausibile, infatti, per risolvere il problema appena citato, nella versione del modello Lee-Carter proposta da Brouhns et al. Nel 2002, i decessi sono distribuiti come una variabile aleatoria di Poisson, cioè:

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

dove:

il termine $E_{x,t}$ rappresenta il numero di esposti al rischio all'età x nell'anno t . esso viene ricavato come la semi-somma dei sopravvissuti di età x al primo Gennaio dell'anno di calendario t , cioè:

$$E_{x,t} = \frac{l_{x,t} + l_{x+1,t}}{2}$$

A questo punto, i parametri κ_t devono essere tali da massimizzare la funzione di log-verosimiglianza rispetto a ζ , ricavata dall'ipotesi che ci sia una distribuzione dei decessi secondo una Poisson, cioè:

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

In alternativa, i parametri κ_t devono essere scelti in maniera tale da far sì che venga minimizzata la devianza:

$$D = 2 \sum_{x=x_1}^{x_m} (D_{x,t} \ln \frac{D_{x,t}}{\hat{D}_{x,t}} - (D_{x,t} - \hat{D}_{x,t}))$$

dove:

- $D_{x,t}$ indica i decessi osservati;
- $\hat{D}_{x,t} = E_{x,t} \exp(\hat{\alpha}_x + \hat{\beta}_x \zeta)$, invece indicano i decessi stimati.

Quindi, rispetto al modello Lee-Carter originale, la metodologia proposta da Brouhns et al. si avvicina maggiormente alla realtà, in quanto assume che la variazione casuale nel numero di decessi segua una distribuzione di Poisson. Questo adattamento rende il modello più adeguato alle età avanzate, poiché è in grado di catturare le fluttuazioni casuali del fenomeno anche quando i dati sono limitati per tali fasce d'età.

2.3.4 Estensione di Lee e Miller (2001)

La variante del modello Lee-Carter proposta da Lee e Miller nel 2001 rappresenta un'estensione e una modifica del modello originale. L'obiettivo principale di tale variante è quello di migliorare la proiezione della mortalità, soprattutto nel lungo termine. Le modifiche proposte da Lee e Miller si concentrano su diversi aspetti, tra cui la ricalibrazione e la proiezione lineare del parametro temporale κ_t . Inoltre, hanno apportato ulteriori cambiamenti che differenziano la loro variante dal metodo Lee-Carter di base.

La valutazione fatta da Lee e Miller si permea di varie fasi:

- **Riduzione del periodo di calibrazione:** Una delle modifiche principali riguarda il periodo di calibrazione del modello, infatti, invece di utilizzare l'intero set di dati storici, Lee e Miller hanno deciso di far iniziare il periodo di calibrazione dal 1950, riducendo così la lunghezza del periodo considerato. Questa scelta è data dal fatto

che le tendenze di mortalità nel tempo sono cambiate significativamente e che i dati più recenti sono più rappresentativi delle tendenze attuali e future.

In particolare, durante la loro valutazione del modello Lee-Carter, Lee e Miller hanno notato che per i dati statunitensi, il modello non performava particolarmente bene se si utilizzava come periodo di riferimento l'intervallo temporale 1900-1989, utilizzato per prevedere il periodo 1990-1997. L'errore principale era dovuto alla discrepanza che c'era tra i tassi di mortalità stimati per l'ultimo anno del periodo di calibrazione, ovvero il 1989, e i tassi effettivi di quell'anno. Tale errore prendeva il nome di "jump-off error" e causava uno scarto di 0.6 anni nell'aspettativa di vita per maschi e femmine combinati. Per evitare tale errore, Lee e Miller hanno vincolato il modello affinché κ_t fosse regolato in modo che passasse attraverso il valore effettivo nell'anno di discontinuità, garantendo una corrispondenza accurata con i tassi di mortalità osservati in quell'anno.

- **Proiezione lineare di κ_t e aggiustamento:** Successivamente, dopo aver ristretto il periodo di calibrazione, Lee e Miller hanno proposto di proiettare κ_t nel futuro utilizzando una regressione lineare basata sulle tendenze recenti. Inoltre, un'altra innovazione è stata quella di effettuare l'aggiustamento di κ_t mediante la calibrazione rispetto all'aspettativa di vita alla nascita nell'anno t . Il contesto al quale si fa riferimento è un contesto nel quale il modello Lee-Carter assume che il pattern di variazione della mortalità sia costante nel tempo. Tuttavia, Lee e Miller hanno notato come il pattern di cambiamento della mortalità, rappresentato dai parametri β_x , variava significativamente tra il periodo 1900-1950 e il periodo 1959-1995. Per riflettere meglio questi cambiamenti, gli autori hanno scelto di utilizzare il 1950 come primo anno del periodo di calibrazione, seguendo una soluzione semplice e soddisfacente, già precedentemente utilizzata da Tuljapurkar et al. nel 2000.
- **Utilizzo dei tassi effettivi nell'anno di discontinuità:** Un'ulteriore modifica riguarda l'uso dei tassi di mortalità effettivi nell'anno di discontinuità, ovvero l'anno nel quale inizia la previsione. Nel modello originale, vi era una differenza tra tassi stimati per l'ultimo anno di calibrazione e quelli effettivi, potando al già citato "jump-off error". Lee e Miller hanno risolto tale problema utilizzando direttamente i tassi effettivi per l'anno di discontinuità, riducendo così l'errore di stima e migliorando la precisione delle proiezioni future.

Le modifiche appena elencate, proposte da Lee e Miller, hanno reso il modello Lee-Carter più flessibile e accurato. Riducendo il periodo di calibrazione, hanno migliorato la rappresentatività del modello rispetto alle tendenze recenti. La proiezione lineare di κ_t e il suo aggiustamento rispetto all'aspettativa di vita alla nascita hanno affrontato il problema della sottostima del declino della mortalità. Infine, l'utilizzo dei tassi effettivi nell'anno di discontinuità ha eliminato l'errore di "jump-off", rendendo le proiezioni più realistiche e affidabili.

Tali modifiche hanno fatto della variante di Lee e Miller un modello più robusto e adattabile, capace di rispondere meglio alle dinamiche della mortalità, soprattutto in contesti in cui le tendenze possono cambiare nel tempo.

2.3.5 Estensione di Booth-Maindonald-Smith (BMS) (2002).

L'estensione proposta da Booth-Maindonald-Smith (BMS) nel 2002 ha come obiettivo quello di migliorare la stabilità e l'accuratezza delle previsioni di mortalità a lungo termine. Tale variante BMS differisce dal modello Lee-Carter sotto 3 aspetti che riguardano la scelta del periodo di calibrazione e la modellazione del parametro temporale κ_t . Le caratteristiche principali della variante BMS sono:

- **Selezione del periodo di calibrazione:** a differenza del modello Lee-Carter originale, dove spesso si utilizza un lungo arco di dati storici, nella variante BMS il periodo di calibrazione viene scelto sulla base di criteri di "goodness-of-fit" statistici, assumendo la linearità di κ_t .
In particolare, in questo caso gli autori hanno applicato il modello Lee-Carter ai dati australiani dal 1907 al 1999, scoprendo che il "modello universale" (Tuljapurkar et al. 2000) del declino costante della mortalità, rappresentato dal parametro κ_t "lineare" non era applicabile a tutto il periodo considerato. Questo ha portato alla decisione di restringere il periodo di calibrazione. Ciò è stato fatto per migliorare l'adattamento del modello, assicurando che l'ipotesi di costanza di β_x , dettata dal modello Lee-Carter sottostante, fosse rispettata.
La scelta del periodo di calibrazione si basa sul rapporto tra le deviazioni medie del modello Lee-Carter sottostante e dell'adattamento lineare complessivo. Tale rapporto è calcolato per tutti i periodi di adattamento (cioè, per tutti gli anni che segnano l'inizio dei periodi, che terminano sempre nell'anno di discontinuità), e viene scelto il periodo per il quale tale rapporto è sostanzialmente inferiore rispetto a quelli dei periodi che iniziano negli anni precedenti.
- **Aggiustamento di κ_t :** un'altra differenza importante, rispetto al modello Lee-Carter, fornita dalla variante BMS è che, mentre nel modello originale κ_t viene solitamente aggiustato in base al numero totale di decessi D_t , nella variante BMS, κ_t viene aggiustato in base alla distribuzione per età dei decessi $D_{x,t}$, utilizzando la distribuzione di Poisson per modellare il processo di morte e la statistica di devianza per misurare la bontà dell'adattamento. Questo approccio consente una migliore rappresentazione delle variazioni nella mortalità specifiche per età.
- **Tassi di "jump off":** in questo caso, i tassi di "jump-off", cioè i tassi di mortalità per l'anno di discontinuità, vengono presi come i tassi stimati (basati sulla metodologia di fitting utilizzata), piuttosto che sui tassi osservati o stimati senza nessun aggiustamento. Questo aiuta a ridurre l'errore di discontinuità ("jump off error") che può influenzare negativamente le proiezioni future.

2.4 Il modello di Renshaw-Haberman. (M2)

Renshaw e Haberman nel 2006, tramite una generalizzazione del modello Lee-Carter, proposero uno dei primi modelli stocastici per la mortalità della popolazione che incorporava un effetto di coorte. Questo fu fatto perché ci si accorse che in alcuni paesi i tassi di mortalità, oltre ad essere influenzati dall'età e dall'anno di calendario, erano influenzati anche dall'anno di nascita della coorte al quale si faceva riferimento. Proprio per questo i due autori hanno ideato tale modello, che si basa sulla seguente relazione:

$$\log m_{x,t} = \beta_x^{(1)} + \beta_x^{(2)} \kappa_t^{(2)} + \beta_x^{(3)} \gamma_{t-x}^{(3)}$$

dove $\kappa_t^{(2)}$ è un effetto di periodo casuale e $\gamma_{t-x}^{(3)}$ è un effetto di coorte casuale ed è funzione dell'anno di nascita (approssimativo) $(t-x)$.

Nella loro analisi, Renshaw e Haberman, hanno scoperto che c'era un miglioramento significativo rispetto al modello Lee-Carter. Infatti, il miglioramento più evidente risiede nel fatto che un'analisi dei residui standardizzati ha rilevato una dipendenza significativamente ridotta dall'anno di nascita, rispetto a quanto osservato con il modello Lee-Carter.

Sfortunatamente, però, il modello Renshaw-Haberman risulta soffrire di una mancanza di robustezza. Cairns et al. (2007, 2008) hanno riscontrato il medesimo problema quando hanno sviluppato il proprio modello. Si ritiene che questa mancanza di robustezza sia legata alla forma della funzione di verosimiglianza. Infatti, per i modelli robusti, la funzione di verosimiglianza ha probabilmente un massimo unico che rimane sostanzialmente invariato se si cambia l'intervallo di anni o di età. Invece, per i modelli nel quale c'è scarsa robustezza, la funzione di verosimiglianza ha probabilmente più di un massimo. Quindi, quando si cambia la fascia di età o l'anno, l'ottimizzatore salterà periodicamente da un massimo all'altro con caratteristiche qualitativamente molto diverse.

Inoltre, Cairns et al. (2008), hanno notato un ulteriore problema del modello di Renshaw e Haberman. Tale problema risiede nel fatto che l'effetto di coorte $\gamma_{t-x}^{(3)}$ sembra avere un andamento deterministico lineare, o forse quadratico, nell'anno di nascita. Ciò suggerisce che l'effetto di coorte viene utilizzato, inavvertitamente, per compensare la mancanza dell'effetto di un secondo periodo di età. Ciò suggerisce che un miglioramento del modello potrebbe essere quello di combinare l'effetto del secondo periodo di età in Renshaw e Haberman con un effetto di coorte più semplice. Ciò potrebbe comportare un adattamento migliore, anche se potrebbe non affrontare il problema della robustezza.

2.5 Il modello di Cairns-Blake-Dowd (2006). (M5)

Il modello di Cairns-Blake-Dowd (CBD), proposto nel 2006, è un modello attuariale utilizzato per descrivere e prevedere l'evoluzione della mortalità nel tempo, facendo riferimento all'andamento delle tavole di mortalità per una popolazione. Tale modello, contraddistinto dalla sigla "M5", appartiene alla classe dei modelli stocastici di mortalità, che cercano di catturare l'incertezza inerente al miglioramento o al peggioramento della longevità nel tempo.

Il modello CBD può essere espresso tramite una relazione molto importante, che è la seguente:

$$(7) \eta_{x,t} = \text{logit}(q_{x,t}) = \ln\left(\frac{q_{x,t}}{1-q_{x,t}}\right) = \beta_x^{(1)} \kappa_t^{(1)} + \beta_x^{(2)} \kappa_t^{(2)}$$

dove: $(1 - q_{x,t}) = p_{x,t}$

Per questo modello sono state assunte semplici forme parametriche per $\beta_x^{(1)}$ e $\beta_x^{(2)}$:

$$\beta_x^{(1)} = 1$$

$$\beta_x^{(2)} = (x - \bar{x})$$

dove, \bar{x} rappresenta l'età media nell'intervallo campione. Allora:

$$(8) \eta_{x,t} = \text{logit}(q_{x,t}) = \ln\left(\frac{q_{x,t}}{1-q_{x,t}}\right) = \kappa_t^{(1)} + \kappa_t^{(2)}(x - \bar{x})$$

il che implica che per un dato anno t , il valore di $q_{x,t}$, dopo una trasformazione logit, è correlato linearmente con l'età x .

Una volta che è stata descritta questa semplice struttura, gli indici di mortalità CBD descritti sopra, ovvero i parametri $\kappa_t^{(1)}$ e $\kappa_t^{(2)}$, possono essere facilmente interpretati.

- Il primo indice di mortalità CBD, $\kappa_t^{(1)}$, rappresenta il livello della curva di mortalità (la curva di $q_{x,t}$ nell'anno t) dopo una trasformazione logit. Uno spostamento parallelo verso il basso della curva di mortalità logit-trasformata rappresenta un miglioramento complessivo della mortalità.

L'impatto di una riduzione di $\kappa_t^{(1)}$ sulla curva di mortalità trasformata è illustrato nel pannello di sinistra della "Figura 7".

- Il secondo indice di mortalità CBD, $\kappa_t^{(2)}$, rappresenta la pendenza della curva di mortalità logit-trasformata.

Un incremento di $\kappa_t^{(2)}$, cioè un aumento della pendenza della curva di mortalità logit-trasformata, sta a significare che la mortalità in età più giovani (al di sotto dell'età media \bar{x}) migliora più rapidamente che in età più avanzata (al di sopra dell'età media \bar{x}).

L'impatto di un aumento di $\kappa_t^{(2)}$ sulla curva di mortalità logit-trasformata è illustrato nel pannello centrale della "Figura 7".

Si noti che, quando il primo indice è fisso, una variazione del secondo indice non ha alcun impatto sulla mortalità all'età media \bar{x} .

I due indici di mortalità del modello CBD possono essere utilizzati per rappresentare una curva di mortalità logit-trasformata con qualsiasi pendenza e livello. Il pannello a destra della "Figura 7" mostra l'impatto sulla curva di mortalità logit-trasformata quando i due indici vengono modificati simultaneamente.

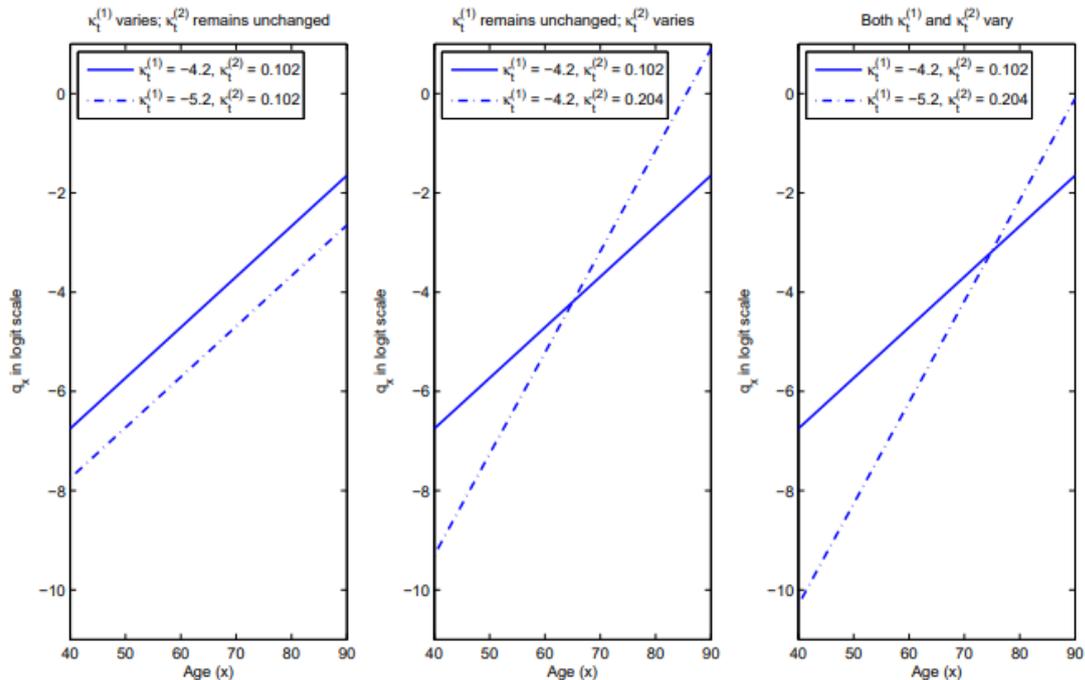


Figura 8: variazioni della curva di mortalità logit-trasformata quando uno o entrambi gli indici di mortalità del CBD cambiano. Fonte: "The CBD Mortality Indexes: Modeling and Applications". Wai-Sum Chan, Johnny Siu-Hang Li and Jackie Li.

Sebbene ogni indice di mortalità del modello CBD abbia il suo significato, è importante considerarli congiuntamente, perché l'associazione tra loro ha un impatto significativo sull'esposizione al rischio di longevità di un portafoglio.

Gli indici di mortalità appena osservati, oltre alle proprietà sopraelencate, hanno anche altri criteri da soddisfare o meno:

- **Inequivocabilità:** la popolazione su cui gli indici di mortalità si basano deve essere definita nel dettaglio. Tale criterio non è difficile da soddisfare se gli indici di mortalità del modello CBD si basano sulle principali popolazioni nazionali, le cui composizioni sono ben dettagliate e documentate.
- **Trasparenza:** il metodo utilizzato per calcolare i valori degli indici di mortalità deve essere chiaro.
- **Oggettività:** il metodo utilizzato per calcolare gli indici di mortalità dovrebbe avere la minor influenza soggettiva possibile. Gli indici di mortalità del modello CBD soddisfano tale criterio in quanto, dato un campione di dati e una fascia di età, la stima della massima verosimiglianza del modello CBD non chiede nessuna informazione soggettiva.
- **Misurabilità:** l'esperienza di mortalità della popolazione al quale si sta facendo riferimento deve poter essere misurabile. Il criterio è soddisfatto se gli indici di mortalità del modello CBD si basano sulle principali popolazioni nazionali.
- **Tempestività e Regolarità:** i dati sulla mortalità per la popolazione di riferimento dovrebbero essere disponibili tempestivamente e aggiornati regolarmente, in maniera tale che i valori gli indici possano essere resi disponibili secondo un

calendario prestabilito. Tale criterio è strettamente legato al fornitore dei dati, infatti, il soddisfacimento di questo criterio richiede una forte collaborazione da parte di tale soggetto.

- **Appropriatezza:** gli indici dovrebbero riflettere la composizione delle popolazioni che richiedono una “copertura”. Se gli indici di mortalità del modello CBD si basano sulle popolazioni nazionali allora tale criterio potrebbe non essere soddisfatto, in quanto, l’esperienza di mortalità di chi richiede una copertura pensionistica potrebbe essere molto diversa dall’esperienza di mortalità della popolazione totale. Questo problema potrebbe essere risolto basando gli indici di mortalità del modello CBD sui dati relativi ai vari piani pensionistici.
- **Numerosità bassa degli indici:** il numero di indici dovrebbe essere basso, in modo tale che i titoli basati su di essi siano liquidi. Gli indici di mortalità del modello CBD soddisfano questo criterio, perché per ogni popolazione di riferimento ci sono solo due indici di mortalità: $\kappa_t^{(1)}$ e $\kappa_t^{(2)}$.
- **Rilevanza:** gli indici di mortalità dovrebbero consentire agli hedger (operatori di mercato che utilizzano strumenti finanziari) di creare coperture efficaci per la longevità.
- **Stabilità e specificabilità anticipata:** gli indici dovrebbero essere stabili e definiti in anticipo per quanto possibile. A meno che la disponibilità dei dati non cambi in futuro, tale criterio è soddisfatto dagli indici di mortalità del modello CBD.

Inoltre, nel corso degli anni numerosi studiosi hanno sviluppato 3 importanti varianti del modello di Cairns-Blake-Dowd, contraddistinte dalle sigle M6-M7-M8.

2.5.1 Il modello Cairns-Blake-Dowd con Effetto Coorte. (M7)

Nel 2009 *Cairns et al.* hanno esteso il modello Cairns-Blake-Dowd originale, aggiungendo un effetto di coorte e un effetto età di ordine quadratico per ottenere il predittore. Tra i diversi modelli analizzati, la seguente generalizzazione del modello a due fattori CBD ha prodotto buoni risultati:

$$\eta_{x,t} = \text{logit}(q_{x,t}) = \ln\left(\frac{q_{x,t}}{1-q_{x,t}}\right) = \kappa_t^{(1)} + \kappa_t^{(2)}(x - \bar{x}) + \kappa_t^{(3)}((x - \bar{x})^2 - \sigma_x^2) + \gamma_{t-x}$$

dove: σ_x^2 rappresenta il valore medio di $(x - \bar{x})^2$.

Rispetto al modello originale CBD, ci sono due componenti aggiuntivi. In primo luogo, c’è un ulteriore effetto del periodo di età, $\kappa_t^{(3)}((x - \bar{x})^2 - \sigma_x^2)$. Tale elemento, seppur sia considerevolmente meno importante dei primi due indici di mortalità, comunque fornisce un miglioramento significativo statisticamente parlando.

In secondo luogo, viene introdotto un effetto di coorte γ_{t-x} , cioè una funzione dell’anno di nascita approssimativo $t-x$.

Questo modello, solitamente indicato con la sigla “M7”, non è identificabile senza le seguenti trasformazioni:

$$\begin{aligned}\kappa_t^{(1)} &\rightarrow \kappa_t^{(1)} + \phi_1 + \phi_2(t - \bar{x}) + \phi_3((t - \bar{x})^2 + \hat{\sigma}_x^2) \\ \kappa_t^{(2)} &\rightarrow \kappa_t^{(2)} - \phi_2 - 2\phi_3(t - \bar{x}) \\ \kappa_t^{(3)} &\rightarrow \kappa_t^{(3)} + \phi_3 \\ \gamma_{t-x} &\rightarrow \gamma_{t-x} - \phi_1 - \phi_2(t - x) - \phi_3(t - x)^2\end{aligned}$$

Si osserva che gli elementi ϕ_1 , ϕ_2 e ϕ_3 sono costanti reali.

Per identificare il modello Cairns et al. (2009) vengono imposti dei vincoli sui parametri molto importanti:

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

tali vincoli assicurano che l'effetto di coorte fluttui intorno allo zero.

Seguendo quanto affermato da Haberman e Renshaw nel 2011, tali vincoli possono essere imposti solo applicando alle trasformazioni, descritte in precedenza, le costanti ϕ_1 , ϕ_2 e ϕ_3 , ottenute facendo regredire γ_{t-x} su $(t - x)$ e su $(t - x)^2$, in modo che:

$$\gamma_{t-x} = \phi_1 + \phi_2(t - x) + \phi_3(t - x)^2 + \epsilon_{t-x}, \quad \text{dove: } \epsilon_{t-x} \sim N(0, \sigma^2) \quad i. i. d.$$

2.5.2 Il modello Cairns-Blake-Dowd con Effetto Coorte. (M6)-(M8)

Inoltre, Cairns et al. (2009) hanno considerato anche strutture predittive più semplici per il predittore $\eta_{x,t}$. Tali strutture sono:

$$\begin{aligned}\eta_{x,t} &= \kappa_t^{(1)} + \kappa_t^{(2)}(x - \bar{x}) + \gamma_{t-x} \\ \eta_{x,t} &= \kappa_t^{(1)} + \kappa_t^{(2)}(x - \bar{x}) + (x_c - x)\gamma_{t-x}\end{aligned}$$

dove: x_c rappresenta un parametro costante da stimare.

Queste strutture sono tipicamente indicate come i modelli M6 e M8, rispettivamente.

Ciascuno dei modelli M6, M7 e M8 è un'estensione del modello M5 con una certa tolleranza per l'effetto di coorte. Di conseguenza, i modelli M5, M6, M7 e M8 possono essere descritti come membri della famiglia dei modelli CBD-Perks generalizzati.

2.6 Il modello Plat.

Il Modello Plat, Plat (2009), mira a combinare il modello CBD (M7) e il modello LC al fine di ottenere un modello appropriato per l'intera fascia di età e per catturare l'effetto coorte, superando così gli svantaggi dei modelli sviluppati precedentemente. Tale modello si basa sulla seguente relazione:

$$\log(m_x) = a_x + \kappa_t^{(1)} + \kappa_t^{(2)}(\bar{x} - x) + \kappa_t^{(3)}(\bar{x} - x)^+ + \gamma_{t-x}$$

dove: $(\bar{x} - x)^+ = \max(\bar{x} - x, 0)$.

Questo modello è ottenuto dalla seguente equazione:

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

Impostando, $\beta_x^{(1)} = 1$, $\beta_x^{(2)} = \bar{x} - x$, $\beta_x^{(3)} = (x - \bar{x})^+$ e $\beta_x^{(0)} = 1$. Inoltre, viene utilizzata la funzione di collegamento logaritmico per mirare al tasso centrale di mortalità.

Il modello Plat è soggetto ai seguenti vincoli:

$$\sum_t \kappa_t^{(1)} = 0, \sum_t \kappa_t^{(2)} = 0, \sum_t \kappa_t^{(3)} = 0, \\ \sum_{c \in C} \gamma_c = 0, \sum_{c \in C} \gamma_c c = 0, \sum_{c \in C} \gamma_c c^2 = 0.$$

Come descritto in Villegas et al. (2015): “i primi tre vincoli assicurano che gli indici di periodo siano centrati intorno allo zero, mentre gli ultimi tre vincoli assicurano che l’effetto coorte fluttui intorno allo zero e non abbia un andamento lineare o quadratico”.

2.7 I criteri di scelta del modello.

Come già detto precedentemente, appartengono ai principali modelli di proiezione della mortalità i seguenti modelli:

Modello	Formula
Lee-Carter (LC)	$\log(m_{x,t}) = \beta_x^{(1)} + \beta_x^{(2)} k_t^{(2)}$
Renshaw-Haberman (RH)	$\log(m_{x,t}) = \beta_x^{(1)} + \beta_x^{(2)} k_t^{(2)} + \beta_x^{(3)} \gamma_{t-x}^{(3)}$
Cairns-Blake-Dowd (CBD)	$\text{logit}(q_{x,t}) = k_t^{(1)} + k_t^{(2)}(x - \bar{x})$
Cairns-Blake-Dowd-2 (CBD-2)	$\text{logit}(q_{x,t}) = k_t^{(1)} + k_t^{(2)}(x - \bar{x}) + \gamma_{t-x}^{(3)}$

Figura 9: Modelli stocastici di mortalità. Fonte: “I modelli di proiezione della mortalità”. Susanna Levantesi e Massimiliano Menzietti.

Inoltre, Cairns et al. (2008), suggeriscono alcuni criteri per scegliere tra i vari modelli. Ad esempio:

- Modello coerente con i dati storici;
- Le dinamiche future a lungo termine del modello devono essere biologicamente ragionevoli;
- Le stime dei parametri devono essere robuste rispetto al periodo di dati e intervalli di età impiegati;
- Le previsioni del modello devono essere robuste rispetto al periodo di dati e intervalli di età impiegati;
- I livelli di previsione dell’incertezza e le traiettorie centrali devono essere plausibili e coerenti con le tendenze storiche e la variabilità dei dati di mortalità;

- Il modello deve essere semplice da attuare mediante metodi analitici o veloci algoritmi numerici;
- Modello relativamente parsimonioso;
- Modello utilizzabile per generare percorsi campione e calcolare intervalli di previsione;
- Modello che consente di integrare l'incertezza del parametro nelle simulazioni;
- Almeno per alcuni Paesi, il modello deve incorporare un effetto stocastico di coorte.

Si osserva che uno dei principali criteri di scelta del modello è quello che si basa sul:

- Bayes Information Criterion (BIC): criterio obiettivo di scelta del modello basato sulla qualità statistica del fit:

$$BIC = l(\hat{\rho}) - 0.5 K \ln(N)$$

dove:

- $\ln(N) \rightarrow$ è la funzione di log-verosimiglianza;
- $\rho \rightarrow$ è l'insieme dei parametri da stimare con la funzione di verosimiglianza;
- $\hat{\rho} \rightarrow$ è la stima di massima verosimiglianza del vettore dei parametri;
- $N \rightarrow$ è il vettore del numero delle osservazioni;
- $K \rightarrow$ è il numero effettivo dei parametri stimati;
- La funzione di log-verosimiglianza del modello è:

$$l(\rho, D, E) = \sum_{x,t} \{ D_{x,t} \ln[E_{x,t} m_{x,t}(\rho)] - E_{x,t} m_{x,t}(\rho) - \ln(D_{x,t}!) \}$$

3 INTELLIGENZA ARTIFICIALE PER LA MODELLAZIONE E PREVISIONE DELLA MORTALITA'.

3.1 Il Machine Learning.

Il Machine Learning, o anche definito apprendimento automatico, è un'applicazione dell'intelligenza artificiale (AI) che fornisce ai sistemi la capacità di apprendere e migliorare automaticamente dall'esperienza senza essere programmati esplicitamente. L'apprendimento automatico si concentra sullo sviluppo di programmi per computer in grado di accedere ai dati e utilizzarli per apprendere da soli. (Tom Mitchell, McGraw Hill, 2997).

Spesso i termini "apprendimento automatico" e "intelligenza artificiale" vengono utilizzati insieme e in modo interscambiabile, ma non hanno lo stesso significato. Un'importante distinzione è che, sebbene tutto ciò che riguarda il machine learning rientra nell'intelligenza artificiale, l'intelligenza artificiale non include solo il machine learning.

Al giorno d'oggi, il machine learning è utilizzato ovunque con molti scopi diversi:

- Effettuare ricerche di mercato e segmentazione: Il pubblico di destinazione è la pietra angolare di qualsiasi business. Ogni impresa ha bisogno di capire il pubblico e il mercato che vuole raggiungere per avere successo. I media e l'industria

dell'intrattenimento usano l'apprendimento automatico per capire cosa piace e cosa non piace al loro pubblico e indirizzare il giusto contenuto a loro.

- Esplorare il comportamento dei clienti: Il machine learning aiuta anche le aziende a esplorare il comportamento del pubblico. Infatti, questo sistema di apprendimento automatico, noto come "user modeling", è il risultato diretto dell'interazione uomo-computer.
- Personalizzare le raccomandazioni: Le imprese hanno bisogno di offrire personalizzazione ai loro clienti. Che si tratti di uno smartphone o di una serie web, le aziende hanno bisogno di stabilire una forte connessione con i loro utenti per fornire ciò che è rilevante per loro.
- Predire le tendenze: Gli algoritmi di apprendimento automatico utilizzano i grandi dati per imparare le tendenze future e prevederle per le aziende. Con l'aiuto di computer interconnessi, una rete di apprendimento automatico può costantemente imparare cose nuove da sola e migliorare le sue capacità analitiche ogni giorno. In questo modo, non si limita a calcolare i dati, ma si comporta come un sistema intelligente che utilizza le esperienze passate per modellare il futuro.
- Aiutare il processo decisionale: L'apprendimento automatico utilizza una tecnica chiamata analisi delle serie temporali che è in grado di analizzare una serie di dati insieme. Esso è un ottimo strumento per aggregare e analizzare i dati e rende più facile per i manager prendere decisioni per il futuro.
- Decodifica dei modelli: Il machine learning può essere altamente efficiente per decifrare i dati nei settori in cui la comprensione dei modelli di consumo può portare a grandi scoperte. Per esempio, settori come quello sanitario e farmaceutico hanno a che fare con un sacco di dati. Il machine learning può aiutarli ad analizzare tutti questi dati per identificare le malattie nella fase iniziale tra i pazienti.

Gli algoritmi sono i motori che alimentano il machine learning. I due tipi principali di algoritmi di machine learning attualmente utilizzati sono gli algoritmi di machine learning "supervisionato" e "non supervisionato". La differenza tra queste due tipologie viene definita dal modo in cui ciascun algoritmo apprende i dati per fare previsioni.

- Modelli "supervisionati": gli algoritmi di machine learning supervisionato sono i più utilizzati. L'obiettivo è quello di prevedere un accadimento futuro e non osservato.
Per farlo vi è una variabile obiettivo osservata y (target) che si vuole prevedere sulla base delle future osservazioni delle variabili esplicative x (features).
Con questo modello, un data scientist agisce da guida e insegna all'algoritmo il risultato atteso (l'algoritmo sa che deve riprodurre Y).
Esempi classici di machine learning supervisionato sono la regressione lineare e logistica, la classificazione multi-classe e gli alberi decisionali.
- Modello "non supervisionati": Il machine learning non supervisionato utilizza un approccio più indipendente, in cui un computer impara a identificare processi e

schemi complessi, senza la guida attenta e costante di una persona.

L'apprendimento automatico senza supervisione implica l'addestramento, basato su dati che non hanno né etichette né un output specifico e definito. Gli algoritmi di clustering k-means, l'analisi di componenti principali e indipendenti e le regole di associazione sono esempi di machine learning non supervisionato.

3.2 Machine e Deep Learning per lo studio della mortalità.

I modelli stocastici di mortalità, visti in precedenza, sono diventati l'approccio standard per la modellazione e la previsione della mortalità. Come già ampiamente detto, il primo contributo pionieristico è stato il modello ideato da Lee e Carter (1992). Da quando tali autori hanno dato il loro contributo, è stata sviluppata un'ampia gamma di estensioni e varianti.

In precedenza, si è osservato come i modelli di mortalità stocastica, in genere, scompongono i tassi di mortalità in tre dimensioni (o in qualsiasi combinazione delle tre): età, periodo e coorte.

Uno dei problemi dei modelli stocastici di mortalità è la loro limitata capacità di rappresentare il pattern dell'andamento della mortalità nei paesi sviluppati, dove il tasso di declino della mortalità ralleza nelle età più giovani e accelera nelle età più avanzate. Questo fenomeno, noto come "rotazione", è stato documentato da *Li et al. (2013)*. Per affrontare questa sfida, *Li et al.* hanno proposto un'estensione del modello Lee-Carter che incorpora la rotazione. Un'altra estensione, proposta da Oeppen (2008), mira a migliorare le previsioni di mortalità prevedendo la distribuzione della tavola di vita dei decessi attraverso l'analisi dei dati composizionali (CoDA) applicata a un modello di tipo LC. I modelli CoDA sono particolarmente efficaci nel prevedere i decessi per causa specifica, poiché tengono conto accuratamente dell'interdipendenza tra le diverse cause di morte.

Mentre i modelli stocastici di mortalità hanno guadagnato terreno nella letteratura sulla mortalità, i metodi di apprendimento automatico basati sui dati dell'informatica e della statistica applicata hanno ricevuto molta attenzione in molti altri campi. Questo sviluppo è stato guidato principalmente dagli enormi progressi nella potenza di calcolo. Nonostante i numerosi vantaggi dell'apprendimento automatico (come la flessibilità, l'evitare di imporre ipotesi strutturali e buone prestazioni all'interno e all'esterno del campione), molti ricercatori all'interno della letteratura sulla mortalità sono ancora riluttanti a utilizzare i metodi. Ciò è dovuto principalmente al fatto che gli algoritmi di apprendimento automatico sono visti come "scatole nere", ovvero che il funzionamento interno degli algoritmi non è chiaro al ricercatore. Di conseguenza, la percezione generale è che i risultati dell'apprendimento automatico "manchino di interpretabilità".

Però, recenti studi all'interno della letteratura sulla mortalità hanno dimostrato che l'apprendimento automatico ha un enorme potenziale nel contesto della modellazione e della previsione della mortalità. Questi studi includono, ad esempio, tecniche basate sugli alberi (ad esempio, Deprez et al., 2017; Levantesi e Pizzorusso, 2019 e Levantesi e Nigri,

2020) e tecniche basate su reti neurali (ad esempio, Hainaut, 2018; Richman e Wuthrich, 2019; Nigri et al., 2019, 2021 e Schnurch e Korn, 2021).

Nella maggior parte degli articoli, l'idea è quella di basarsi sui modelli di mortalità esistenti aggiungendo un livello di apprendimento automatico, migliorando così l'adattamento e le previsioni prodotte da questi modelli. In particolare, *Deprez et al. (2017)* testano retrospettivamente i modelli Lee-Carter e Renshaw e Haberman, utilizzando un metodo di potenziamento degli alberi di regressione di Poisson per calibrare un parametro, che viene poi impiegato per costruire tassi di mortalità migliorati.

Infatti, sulla base di questa idea Levantesi e Pizzorusso nel 2019 hanno dimostrato come l'adattamento e la previsione del modello Lee-Carter possono essere migliorati utilizzando tecniche di machine learning come, ad esempio, i "regression trees", i "random forest" o il "gradient boosting". Invece, in un lavoro successivo, Levantesi e Nigri (2020) estendono il lavoro citato in precedenza combinando i random forest con delle P-spline bidimensionali. L'obiettivo di tale operazione è quello di migliorare l'adattamento e la previsione del modello Lee-Carter.

Inoltre, il lavoro *Nigri et al. (2019)*, per la previsione della componente temporale del modello LC, propone un'alternativa all'approccio ARIMA proposto da Box-Jenkins. In particolare, gli autori suggeriscono di integrare il modello LC con un sistema di "Deep Learning", utilizzando una rete neurale ricorrente. Schnurch e Korn (2021) propongono l'uso di reti neurali convoluzionali per modellare e prevedere la mortalità, implementando una tecnica basata sul bootstrap per quantificare l'incertezza delle previsioni delle reti neurali.

I vantaggi delle tecniche basate su "alberi" rispetto alle "reti neurali" sono che sono meno costose dal punto di vista computazionale e relativamente facili da capire e applicare, questo perché gli alberi possono essere resi "interpretabili". Questo è probabilmente il vantaggio più importante delle tecniche basate sugli alberi, poiché l'"interpretabilità" è spesso considerata più importante della previsione delle prestazioni.

3.2.1 Applicazione delle tecniche di Machine Learning.

Sulla base dell'idea proposta da *Deprez et al. (2017)*, S.Levantesi e V.Pizzorusso nel 2019, hanno proposto un articolo intitolato "Application of Machine Learning to Mortality Modeling and Forecasting", nel quale viene approfondito l'utilizzo di algoritmi di machine learning, in grado di catturare modelli che non sono comunemente identificabili, migliorando la bontà dell'adattamento dei modelli standard di mortalità stocastica. Gli algoritmi di machine learning maggiormente utilizzati in questo caso sono:

- Decision Trees (alberi decisionali)
- Random forest (foresta casuale)
- Gradient Boosting (potenziamento del gradiente)

In particolare, le due autrici, all'interno del loro articolo, mostrano come l'implementazione e l'utilizzo di tali tecniche di apprendimento automatico, che sono basate su componenti caratteristiche come età, sesso, anno di calendario e coorte di nascita, porti ad un migliore adattamento dei dati storici rispetto alle stime fornite dai comuni modelli stocastici di mortalità.

3.2.1.1 Il modello.

La prima cosa da fare è quella di descrivere il modello al quale si fa riferimento nel quale vengono considerate le seguenti variabili categoriali, che identificano un individuo:

- Sesso (g)
- Età (a)
- Anno solare (t)
- Anno di nascita (c)

Successivamente ad ogni individuo viene assegnata la caratteristica $x = (g, a, t, c) \in X$ con

$$X = G \times A \times T \times C$$

che rappresenta lo spazio delle caratteristiche, dove:

- $G = \{\text{maschi}, \text{femmine}\}$;
- $A = \{0, \dots, \omega\}$;
- $T = \{t_1, \dots, t_n\}$;
- $C = \{c_1, \dots, c_m\}$.

Anche altre variabili categoriali potrebbero essere incluse, come ad esempio lo stato civile, il reddito o comunque qualsiasi informazione individuale.

Successivamente, si assume che il numero di decessi D_x soddisfi le seguenti condizioni:

- D_x sono indipendenti in $\{x \in X\}$;
- $D_x \sim \text{Pois}(m_x \cdot E_x)$ per tutti $\{x \in X\}$.

dove, m_x è il tasso centrale di mortalità ed E_x sono gli esposti al rischio.

Inoltre, si indica con d_x^{mdl} il numero atteso di decessi stimato da un modello standard di mortalità stocastica (come Lee-Carter, Cairns-Blake-Dowd, ecc.) e con m_x^{mdl} il corrispondente tasso centrale di mortalità.

Modellando il tasso centrale di mortalità (m_x) invece del tasso di mortalità (q_x), si assume che:

- $m_x = m_x^{mdl}$
- $D_x \sim \text{Pois}(\psi_x \cdot d_x^{mdl})$, con $\psi_x \equiv 1$, $d_x^{mdl} = m_x^{mdl} E_x$

La condizione $\psi_x \equiv 1$ significa che il modello di mortalità specificato si adatta perfettamente ai tassi grezzi. Tuttavia, nel mondo reale, un modello di mortalità potrebbe sovrastimare ($\psi_x \leq 1$) o sottostimare ($\psi_x \geq 1$) i tassi grezzi.

Pertanto, viene calibrato il parametro ψ_x , in base alla caratteristica x , secondo tre diverse tecniche di apprendimento automatico. Si trova ψ_x come soluzione di un “regression tree algorithm” applicato al rapporto tra le osservazioni di morte e il corrispondente valore stimato dal modello di mortalità $\frac{D_x}{d_x^{mdl}}$:

$$(9) \quad \frac{D_x}{d_x^{mdl}} \sim \text{sex} + \text{age} + \text{year of birth} + \text{year of death}$$

Successivamente, si indica con $\hat{\psi}_x^{mdl,ML}$ lo stimatore di Machine Learning ottenuto risolvendo l’equazione (9), dove l’apice *mdl* indica il modello di mortalità stocastica (come, ad esempio, il modello Lee-Carter) e l’apice *ML* indica l’algoritmo di Machine Learning utilizzato per migliorare i tassi di mortalità dati da un certo modello.

Lo stimatore $\hat{\psi}_x^{mdl,ML}$ viene quindi applicato al tasso centrale di mortalità del modello di mortalità specificato, m_x^{mdl} , con l’obiettivo di ottenere un migliore adattamento dei dati osservati:

$$(10) \quad m_x^{mdl,ML} = \hat{\psi}_x^{mdl,ML} \cdot m_x^{mdl}, \quad \forall x \in X$$

Così come avviene in *Deprez et al. (2017)*, viene misurato il miglioramento dei tassi di mortalità raggiunto dall’algoritmo:

$$\Delta m_x^{mdl,ML} = \frac{m_x^{mdl,ML} - m_x^{mdl}}{m_x^{mdl}} = \hat{\psi}_x^{ML} - 1$$

Per stimare $\hat{\psi}_x^{mdl,ML}$, vengono utilizzate le seguenti tecniche di Machine Learning (ML) basate su alberi:

- Decision Tree
- Random Forest
- Gradient Boosting

3.2.1.2 Decision Tree.

I metodi ad albero per la regressione e la classificazione sono diventati alternative popolari alla regressione lineare (*Breiman et al. 1984*). Nell’accezione moderna vengono chiamati anche Classification and Regression Trees (CART) ed essi vengono definiti come degli algoritmi di apprendimento supervisionati utilizzati principalmente in processi decisionali non lineari per risolvere problemi di classificazione o regressione. Questi metodi dividono lo spazio delle caratteristiche X attraverso una serie di divisioni binarie, e l’insieme delle regole di suddivisione può essere rappresentato graficamente come un albero (tree). Dopo aver suddiviso ricorsivamente lo spazio delle caratteristiche in diverse regioni semplici, la risposta per una specifica osservazione viene prevista calcolando la media delle osservazioni di addestramento nella regione di riferimento. (*James et al., 2017*).

Indicando con $(X_\tau)_{\tau \in T}$ la partizione di X ; lo stimatore Decision Tree (DT) viene calcolato come segue:

$$\hat{\psi}(x) = \sum_{\tau \in T} \hat{\psi}_\tau \mathbb{1}_{\{x \in X_\tau\}}$$

dove: $\mathbb{1}_{\{ \}}$ rappresenta la funzione indicatrice.

Le regioni $\{X\tau\}_{\tau \in T}$ sono identificate minimizzando la somma dei quadrati degli scarti (residual sum of squares, RSS):

$$\sum_{\tau \in T} \sum_{i \in X\tau} (\psi_i - \hat{\psi}_{X\tau})^2$$

La variabile target $\hat{\psi}_{X\tau}$ è stimata dalla media (la moda nei problemi di classificazione) delle variabili appartenenti alla medesima regione $X\tau$.

Gli algoritmi di Decision Trees (DT) offrono alcuni vantaggi rispetto ad altri modelli di regressione. Come evidenziato da *James et al. (2017)*, tali algoritmi sono facili da interpretare, gestiscono agevolmente predittori qualitativi senza la necessità di creare variabili fittizie, e hanno la capacità di cogliere ogni tipo di correlazione nei dati essendo totalmente non parametrici.

Tuttavia, presentano anche alcuni svantaggi significativi: risultano essere meno competitivi rispetto ad altri algoritmi di ML supervisionati per qualità delle predizioni e inoltre possono mancare di robustezza poiché piccole modifiche nei dati possono portare alla stima di alberi molto diversi da quelli iniziali. Un altro svantaggio importante è che il dataset utilizzato deve essere piuttosto grande per costruire buoni predittori. Per ovviare a tali svantaggi sono stati proposti i metodi "ensemble" che possono portare notevoli miglioramenti in termini di accuratezza della predizione, a svantaggio di una parziale perdita di interpretabilità del modello. L'obiettivo principale dei metodi "ensemble" è combinare le previsioni di stimatori differenti allo scopo di migliorare la capacità di generalizzazione e la robustezza che si ottiene con un singolo stimatore. Di metodi "ensemble" ne esistono diverse tipologie, tra i quali è possibile identificare i:

- Metodi avarage: costruiscono i differenti stimatori in modo indipendente e calcolano la media delle loro previsioni. In media, lo stimatore ensemble è spesso migliore di qualsiasi stimatore singolo, poiché presenta una minore varianza (es. bagging, random forest).
- Metodi boosting: costruiscono stimatori di base in modo sequenziale con l'obiettivo di ottenere una riduzione del bias. Lo stimatore ensemble si ottiene come combinazione di differenti stimatori deboli (es. gradient boosting).

3.2.1.3 Random Forest.

I Random Forest rappresentano quella tecnica di ML che consiste nell'aggregazione di diversi decision trees, ottenuti generando campioni di bootstrap da dati di training (Breiman, 2001), dando origine ad una "foresta". Siccome si è osservato che l'aggregazione di molti alberi decisionali può migliorare le prestazioni predittive degli alberi, viene applicato in primis il "bagging" (chiamato anche aggregazione bootstrap) per generare un certo numero, B , di decision trees a partire da campioni di addestramento bootstrap, ottenuti tramite il bootstrap del set di dati di addestramento originale. Il Random forest (RF) differisce dal bagging per il modo in cui vengono considerati i predittori: negli

algoritmi RF, ad ogni split dell'albero, viene preso in considerazione sono un sottoinsieme casuale di predittori, come spiegato in dettaglio da Breiman (2001). In tal modo previene la predominanza di predittori forti nelle suddivisioni di ogni albero. Lo stimatore RF viene quindi calcolato come segue:

$$\hat{\psi}(x) = \frac{1}{B} \sum_{b=1}^B \hat{\psi}^{(b)}(x)$$

dove: B rappresenta il numero di campioni bootstrap. Mentre $\hat{\psi}^{(b)}(x)$ rappresenta lo stimatore DT sul campione b .

Lo stimatore RF è ottenuto applicando l'algoritmo del pacchetto R "randomForest" (Liaw 2018). Poiché questa procedura si è rivelata molto costosa da un punto di vista computazionale, il numero di alberi deve essere scelto con cura: non deve essere troppo grande, ma allo stesso tempo deve essere in grado di produrre un'adeguata percentuale di varianza spiegata. Un altro requisito importante è che bisogna ottenere la più bassa media dei residui al quadrato (MSR: Mean Square Residuals).

Inoltre, il numero di alberi nella foresta deve essere sufficientemente grande in modo che ogni predittore abbia abbastanza possibilità di essere selezionato, sebbene un numero relativamente piccolo di alberi (poche centinaia) potrebbe essere già sufficiente per raggiungere un'elevata precisione.

Quindi, l'algoritmo RF consente una riduzione della varianza, ottenuta riducendo la correlazione tra gli alberi (nel processo di crescita degli alberi la correlazione viene ridotta con la selezione casuale delle variabili in input).

Un predittore molto forte all'interno del dataset potrebbe non essere considerato, lasciando spazio agli altri predittori che avranno maggiori possibilità di essere scelti come candidati per l'operazione di suddivisione dell'albero.

Inoltre, così come i decision tree, anche i RF hanno dei vantaggi e degli svantaggi. Tra i vantaggi identifichiamo:

- Pochi parametri;
- Scarsa probabilità di overfitting, se il numero di campioni è sufficientemente alto;
- Migliora la stabilità dei decision trees (grazie al bagging) e ne riduce l'overfitting;
- Riduce la varianza del modello senza peggiorare sensibilmente il bias.

Invece, gli svantaggi sono:

- Si perde l'immediata comprensibilità;
- La fase di prediction può essere molto lenta.

3.2.1.4 Gradient Boosting.

Il Gradient Boosting (GB), anche definito potenziamento del gradiente, è un algoritmo ad albero proposto da Friedman (2001) che utilizza alberi decisionali di dimensioni fisse come modelli predittivi deboli (tipicamente alberi con un numero piccolo di split).

Considerando la perdita associata all'uso di una specifica funzione per prevedere una variabile sui dati di addestramento, il GB cerca di minimizzare tale perdita all'interno del campione. Lo fa attraverso un algoritmo di apprendimento adattivo a stadi che combina predittori deboli per migliorare le prestazioni complessive.

Sia $\psi(x)$ la funzione; l'algoritmo di gradient boosting trova un'approssimazione, indicata con $\hat{\psi}(x)$, alla funzione $\psi(x)$ che minimizza il valore atteso della funzione di perdita differenziabile specificata (problema di ottimizzazione).

Ad ogni stadio i del gradient boosting ($1 \leq i \leq N$), supponiamo che ci siano alcuni modelli imperfetti $\hat{\psi}(x_i)$, quindi l'algoritmo di GB migliora su $\hat{\psi}(x_i)$ costruendo un nuovo modello che aggiunge uno stimatore h per fornire un modello migliore:

$$\hat{\psi}(x_i) = \hat{\psi}(x_{i-1}) + \lambda_i h_i(x)$$

dove: $h_i \in H$ rappresenta una funzione di apprendimento di base e λ è un moltiplicatore ottenuto risolvendo il problema dell'ottimizzazione.

Lo stimatore GB viene ottenuto utilizzando il pacchetto R `gbm` (Ridgeway, 2007). Questo pacchetto richiede la configurazione di diversi parametri chiave, tra cui il numero di alberi (`n.trees`), il numero di ripetizioni per la convalida incrociata (`cv.folds`), la profondità di ogni albero (`interaction.depth`), e il tasso di apprendimento (`shrinkage`). È fondamentale selezionare con attenzione il numero di alberi, poiché un numero elevato può ridurre l'errore sul set di addestramento, mentre un numero troppo basso può causare overfitting. Inoltre, la scelta del numero di ripetizioni per la convalida incrociata dipende dalle dimensioni del set di dati. In genere, una convalida incrociata a cinque ripetizioni, che utilizza il 20% dei dati per i test, è considerata una buona pratica. Inoltre, la profondità di interazione indica il numero massimo di nodi per albero o il livello più alto di interazioni tra variabili consentito nel modello.

3.2.1.5 Fitting del modello.

All'interno del loro lavoro, Levantesi e Pizzorusso (2019), hanno montato i modelli Lee-Carter, Renshaw-Haberman e Plat sulla popolazione italiana. Nell'analisi sono stati presi in considerazione i seguenti insiemi di genere, età, anni e coorte:

- $G = \{\text{maschi}, \text{femmine}\};$
- $A = \{0, \dots, 100\};$
- $T = \{1915, \dots, 2014\};$
- $C = \{1815, \dots, 2014\}$

L'accuratezza del modello viene misurata in base a 2 criteri proposti da Bayes e Akaike che sono rispettivamente il BIC e l'AIC. Ciò che si è osservato è che il modello di Renshaw e

Haberman si adatta molto bene ai dati storici. Tale modello è poi seguito in ordine dal modello Lee-Carter e infine dal modello Plat.

Inoltre, la bontà dell'adattamento è testata anche tramite un'analisi dei residui. Anche in questo caso quello che si osserva è che il modello RH ha fornito un miglior adattamento, nonostante il maggior numero di parametri. Invece, il modello LC ha fornito un buon adattamento, soprattutto per la popolazione anziana, mentre il modello Plat ha fornito le prestazioni peggiori, nonostante l'elevato numero di parametri coinvolti.

Successivamente, si osserva come il fitting del modello può essere migliorato utilizzando tecniche di machine learning. In particolare, i parametri utilizzati per calibrare gli algoritmi di machine learning descritti in precedenza sono i pacchetti "rpart", "randomForest" e "gbm":

- $\hat{\psi}_x^{mdl,DT}$ è stimato con il pacchetto "rpart" impostando $cp = 0,003$ (parametro di complessità);
- $\hat{\psi}_x^{mdl,RF}$ è stimato con il pacchetto "randomForest" impostando: $ntrees = 200$ (numero di alberi). Poiché la procedura di Random Forest si è rivelata molto costosa da un punto di vista computazionale, si è preferito limitare il numero di alberi a 200, in modo da garantire, oltre ad un'adeguata percentuale di varianza spiegata dal modello, anche una bassa media di quadrati residui, MSR;
- $\hat{\psi}_x^{mdl,GB}$ è stimato con il pacchetto "gbm" impostando: $n.trees = 5000$ (numero di alberi); $cv.folds = 5$ (numero di pieghe di convalida incrociata); $interaction.depth = 6$; $shrinkage = 0.001$ (tasso di apprendimento). Il parametro $cv.folds$ viene utilizzato per stimare il numero ottimale di iterazioni.

Il livello di miglioramento dei tassi centrali di mortalità derivante dall'applicazione di algoritmi di machine learning è stato misurato dal parametro $\Delta m_x^{mdl,ML}$. I risultati numerici per il modello LC, RH e Plat combinati con gli algoritmi di ML basati su alberi sono mostrati nella seguente figura:

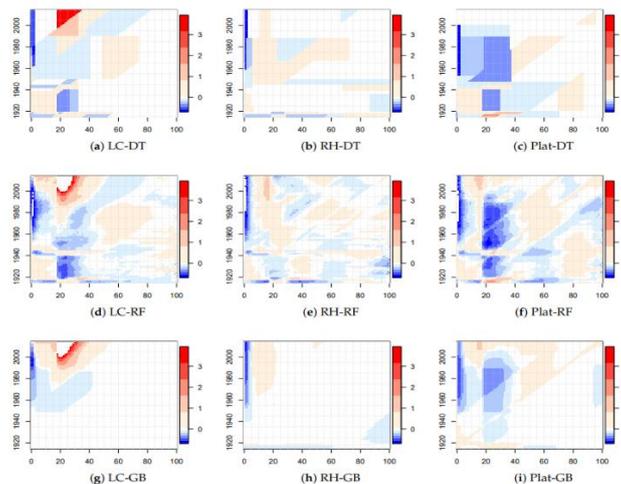


Figura 10: Risultati numerici per il modello LC, RH e Plat combinati con gli algoritmi di ML basati su alberi. Fonte: S.Levantesi e V. Pizzorusso: "Application of Machine Learning to Mortality Modeling and Forecasting".

Questa figura ci fornisce una rappresentazione dei valori di $\Delta m_x^{mdl,ML}$, per quanto riguarda la popolazione maschile italiana (Età 0-100 e anni 1915-2014). I risultati per le donne sono molto simili.

In questa rappresentazione, le aree bianche rappresentano le variazioni molto piccole di $\Delta m_x^{mdl,ML}$, approssimativamente intorno allo zero. Aree bianche più grandi sono state osservate per l'aumento del gradiente applicato al modello LC e al modello RH. Comunque, in tutti i casi ci sono stati anche cambiamenti significativi. Tali cambiamenti risultano essere meno evidenti per il modello RH che, come ben sappiamo meglio si adattava ai dati storici.

Inoltre, da queste rappresentazioni si osserva anche che molte regioni sono caratterizzate da divisioni diagonali, evidenziando un effetto di coorte. Ciò rafforza la scelta di inserire il parametro di coorte negli algoritmi di decision tree.

Successivamente si osserva che, soprattutto per il modello Lee-Carter, ci sono stati cambiamenti che si sono concentrati soprattutto nelle età giovanili.

Invece, per il modello Plat sono stati osservati piccoli valori di $\Delta m_x^{mdl,ML}$ rispetto agli alti modelli stocastici di mortalità, con l'eccezione della popolazione di età inferiore ai 40 anni che ha mostrato variazioni piuttosto significative.

Da questi primi risultati, il decision tree e il random forest sembrano funzionare meglio rispetto al gradient boosting.

Allora, analizzando tali grafici, è possibile sostenere che gli stimatori di ML hanno portato a un miglioramento della qualità dell'adattamento in tutti i modelli di mortalità considerati. I grafici mostrano che l'applicazione di uno stimatore di ML comporta cambiamenti significativi nei valori del "tasso di mortalità" con un miglioramento significativo nel fitting dei dati. Tra i modelli considerati, il modello Plat è quello che ha ottenuto maggiori miglioramenti del fitting, grazie all'uso di algoritmi di ML.

Successivamente, viene misurata la bontà dell'adattamento dei modelli attraverso un indicatore molto importante chiamato MAPE (mean absolute percent error). Tale indicatore è definito come segue:

$$MAPE = \frac{100}{N} \sum_x \left| \frac{m_x^{mdl,ML} - \hat{m}_x^{mdl,ML}}{m_x^{mdl,ML}} \right|$$

dove N è la dimensione dei dati; $m_x^{mdl,ML}$ e $\hat{m}_x^{mdl,ML}$ sono rispettivamente i valori "effettivi" e "stimati" della mortalità. I risultati del MAPE sono riassunti nella seguente tabella:

Gender	Males			Females		
	LC	RH	Plat	LC	RH	Plat
No ML	19.48%	18.03%	25.81%	13.83%	13.96%	22.34%
DT	11.06%	11.57%	13.69%	9.42%	9.12%	11.53%
RF	4.85%	4.89%	4.79%	4.61%	4.65%	4.49%
GB	13.80%	11.35%	14.59%	8.48%	7.84%	9.77%

Figura 11: Valori del MAPE per i tre modelli (LC, RH e Plat) prima e dopo l'utilizzo del ML. Fonte: S.Levantesi e V. Pizzorusso: "Application of Machine Learning to Mortality Modeling and Forecasting"

Questa tabella indica i valori del MAPE per i tre modelli (LC, RH e Plat) prima dell'utilizzo del ML e dopo l'utilizzo delle 3 tecniche del ML.

La riduzione più alta del MAPE è stata evidenziata dal modello Plat, con una riduzione dal 25,81% al 4,79% dopo l'applicazione dell'algoritmo RF (dal 22,34% al 4,49% per le donne). In sintesi, tutti gli algoritmi di machine learning hanno migliorato i modelli standard di mortalità stocastica considerati. Infine, si osserva come il Random Forest si è rivelato l'algoritmo più efficace.

3.2.1.6 Previsione del modello Lee-Carter migliorata dal Machine Learning.

Riprendendo i modelli stocastici osservati nel capitolo due, in questa sottosezione, viene descritto come lo stimatore di Machine Learning $\hat{\psi}_x$, può essere utilizzato per ottenere un miglioramento della previsione di mortalità data dai modelli stocastici standard, come ad esempio il modello LC.

Mettendo un attimo da parte la logica che accompagna l'apprendimento automatico, l'idea alla base proposta da Levantesi e Pizzorusso (2019), era quella di modellare e prevedere $\hat{\psi}_x$ utilizzando lo stesso quadro del modello di mortalità originale. Dunque, i valori previsti di $\hat{\psi}_x$ sono stati utilizzati per migliorare i valori previsti dei tassi di mortalità ottenuti mediante il modello originale.

Nell'esempio in questione, tale approccio è testato sul modello LC; pertanto, lo stimatore di ML $\hat{\psi}_x^{LC,ML}$ è modellato come segue:

$$(11) \quad \log(\hat{\psi}_x^{LC,ML}) = \alpha_x^\psi + \beta_x^{(1,\psi)} \kappa_t^{(1,\psi)}$$

dove gli insiemi di parametri $\alpha_x^\psi, \beta_x^{(1,\psi)}, \kappa_t^{(1,\psi)}$ hanno lo stesso significato di $\alpha_x, \beta_x, \kappa_t$ nell'equazione 1 (LC model).

Successivamente si osserva che combinando l'equazione (1), l'equazione (10) e l'equazione (11), otteniamo il Modello Lee-Carter migliorato dall'algoritmo di machine learning:

$$(12) \quad \log(m_x^{LC,ML}) = (\alpha_x^\psi + a_x) + \beta_x^{(1,\psi)} \kappa_t^{(1,\psi)} + \beta_x \kappa_t$$

Per verificare l'accuratezza del modello vengono fornite delle previsioni fuori campione, in cui il periodo di adattamento fa riferimento all'intervallo di tempo 1915-2000, mentre il periodo di previsione fa riferimento all'intervallo 2001-2014. Durante la fase di previsione,

$\kappa_t^{(1,\psi)}$ e κ_t vengono entrambi modellati da un “random walk with drift”, utilizzando i dati degli ultimi 40 anni (1960-2000).

I valori del parametro κ_t che fa riferimento al modello LC, sono stati fortemente decrescenti dalla fine della Seconda guerra mondiale, il che ha comportato una forte riduzione della mortalità nel tempo, con un’ulteriore accelerazione dopo la metà degli anni ’80. Gli algoritmi di ML hanno ridotto questo effetto attraverso il parametro $\kappa_t^{(1,\psi)}$ che ha mostrato una tendenza alla crescita dopo il 1960, con maggiore forza a partire dagli anni ’80. Ciò lo si può vedere nella seguente figura:

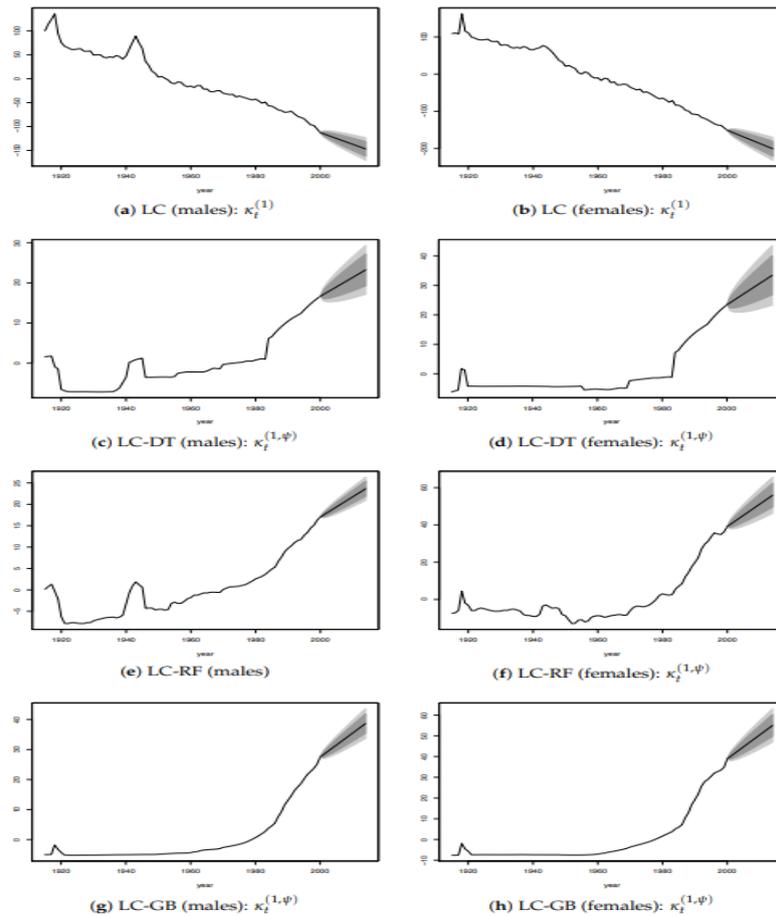


Figura 12: Rappresentazione dei valori di $\kappa_t^{(1,\psi)}$ e κ_t , pre e post utilizzo del ML. Fonte: S.Levantesi e V. Pizzorusso: “Application of Machine Learning to Mortality Modeling and Forecasting”

Questa è la rappresentazione dei valori di $\kappa_t^{(1,\psi)}$ e κ_t , pre e post utilizzo del ML, sia per gli uomini che per le donne, considerando “fitted value” (1915-2000) e “forecasted value” (2001-2014).

Gli autori sottolineano che l’adozione del modello di mortalità originale sia per l’adattamento che per la previsione dello stimatore di ML porta a un miglioramento della proiezione e rende facile comprendere l’effetto di tali miglioramenti sui parametri del modello.

Come menzionato in precedenza, l'apprendimento automatico è noto per la sua efficacia nel rilevare schemi sconosciuti e difficili da identificare nei dati, ma è privo di una solida teoria di base che possa offrire una spiegazione razionale dei risultati ottenuti. In questo contesto, l'approccio appena descritto mira a colmare il divario tra il machine learning e la teoria, combinando un metodo guidato dai dati con uno basato sui modelli.

A questo punto si va a valutare la bontà delle previsioni. Ciò lo si fa utilizzando due misure che sono:

- Il "Root Mean Squared Logarithmic Error" (RMSLE);
- Il "Root Mean Squared Error" (RMSE).

Il primo (RMSLE) tiene conto del $\log(m_x)$, attribuendo un peso maggiore agli errori nelle età più giovani, mentre il secondo (RMSE) si basa su (m_x) e dà un peso più significativo agli errori nelle età più avanzate.

$$RMSLE = \sqrt{\frac{\sum_x (\log \hat{m}_x - \log m_x)^2}{N}}$$

$$RMSE = \sqrt{\frac{\sum_x (\hat{m}_x - m_x)^2}{N}}$$

La seguente tabella mostra i valori del RMSLE e del RMSE del modello LC, pre e post utilizzo del ML:

Model	RMSLE		RMSE	
	Males	Females	Males	Females
LC	0.0290	0.0195	0.7282	0.7734
LC, DT	0.0139	0.0068	0.3567	0.5351
LC, RF	0.0083	0.0044	0.3624	0.1532
LC, GB	0.0100	0.0056	0.3536	0.1841

Figura 13: Valori del RMSLE e del RMSE del modello LC, pre e post utilizzo del ML. Fonte: S. Levantesi e V. Pizzorusso: "Application of Machine Learning to Mortality Modeling and Forecasting"

I valori in grassetto indicano il modello con i valori RMSLE e RMSE più piccoli. Ciò che si osserva è che l'algoritmo RF ha fornito le prestazioni migliori, ad eccezione del RMSE per i maschi, dove il gradient boosting è stato il migliore. La riduzione più significativa dell'RMSLE è stata del 77% per gli uomini e del 71% per le donne, mentre per l'RMSE, la riduzione è stata del 51% per gli uomini e dell'80% per le donne. In ogni caso si può affermare che tutti gli stimatori basati sul machine learning hanno portato a un miglioramento sostanziale delle previsioni rispetto al modello Lee-Carter standard.

A questo punto, si va ad analizzare il livello di miglioramento fornito dagli algoritmi di machine learning, però non più sul set di dati completo, ma su un set di dati più piccolo,

ovvero quello che fa riferimento all'intervallo di tempo 1960-1914. L'obiettivo è quello di verificare se la modifica del periodo di calibrazione (con inizio che passa da 1915→1960) può avere un impatto importante sui risultati, poiché gli algoritmi di ML funzionano meglio con set di dati più grandi. Nella seguente tabella sono riportati i valori del MAPE, pre e post utilizzo degli algoritmi di ML, utilizzati per analizzare la qualità del fitting. Tale tabella adesso, al contrario di quella vista in precedenza fa riferimento all'intervallo 1960-2014, ed è la seguente:

Gender	Males			Females		
	Model	LC	RH	Plat	LC	RH
No ML	11.08%	8.51%	12.67%	9.20%	8.44%	12.32%
DT	5.86%	5.27%	6.25%	5.90%	5.71%	6.62%
RF	4.07%	4.02%	3.89%	4.95%	4.79%	4.81%
GB	6.04%	4.84%	6.07%	5.49%	5.16%	6.27%

Figura 14: Valori del MAPE, pre e post utilizzo degli algoritmi di ML, utilizzati per analizzare la qualità del fitting. Fonte: S.Levantesi e V. Pizzorusso: "Application of Machine Learning to Mortality Modeling and Forecasting"

Ciò che si osserva è che, per un periodo di calibrazione più breve, tutti gli algoritmi di ML hanno migliorato i livelli standard di mortalità stocastica e il livello di miglioramento del fitting è rimasto elevato. Il random forest ha continuato ad essere il migliore.

Successivamente, sempre per l'intervallo 1960-2014, si osservano i valori di $\Delta m_x^{mdl,ML}$, per i modelli LC, RH e Plat combinati con gli algoritmi di machine learning. I risultati sono molto simili sia per gli uomini che per le donne e anche in questo caso ci sono stati cambiamenti significativi, ma ci sono state meno regioni identificate da divisioni diagonali rispetto al periodo 1915-2014. Inoltre, si osserva come tali cambiamenti significativi si sono concentrati sia in età giovani che in età anziane. La concentrazione in età anziane era meno evidente rispetto all'intervallo 1915-2014.

3.2.1.7 Considerazioni finali.

Tramite tutte queste procedure, è stato dimostrato come l'apprendimento automatico possa potenziare l'adattamento e la previsione di modelli di mortalità stocastica tradizionali come il LC, RH e il Plat. Viene utilizzata l'intelligenza artificiale, del quale fa parte il ML, per approfondire i processi che rimangono invisibili ai modelli standard. Gli algoritmi utilizzati sono stati il decision tree, il random forest e il gradient boosting e i risultati evidenziano come l'approccio basato sul RF risulta essere quello più efficace, pur confermando miglioramenti significativi anche con gli altri due approcci.

Un'altra novità che è stata proposta da Levantesi e Pizzorusso (2019), è l'introduzione di uno stimatore di ML per migliorare la qualità delle previsioni fornita dai modelli di mortalità standard, in cui l'apprendimento automatico è stato utilizzato come supporto e

non come sostituto ad essi. Lo stimatore di ML è stato previsto utilizzando il medesimo framework del modello di mortalità originale. Questo approccio, non solo mira a migliorare i tassi di mortalità previsti dal modello LC standard, ma anche a creare un collegamento tra l'apprendimento automatico e la teoria, facilitando una spiegazione razionale dei risultati ottenuti.

In conclusione, si può affermare che l'utilizzo delle tecniche di machine learning sopra elencate migliora significativamente la qualità della stima della mortalità, soprattutto quando queste tecniche sono combinate con i modelli di mortalità stocastica standard.

3.2.2 Applicazione delle tecniche di Deep Learning.

L'articolo intitolato "A Deep Learning Integrated Lee-Carter Model", proposto da Nigri et al. nel 2019, fornisce un approccio alternativo al processo ARIMA che viene solitamente utilizzato per modellare il parametro κ_t all'interno del modello Lee-Carter. Ciò viene fatto perché si osserva che il processo ARIMA mostra evidenti limiti nel descrivere la mortalità futura.

Allora, l'approccio alternativo che viene proposto dagli autori è basato su una tecnica di Deep Learning. Il termine Deep Learning sta a significare "apprendimento profondo" ed esso prende spunto dalla struttura del cervello umano e dall'interconnessione tra i neuroni. In particolare, serve a creare modelli di apprendimento che si basano su più livelli (ad esempio le reti neurali multistrato) che combinano diversi algoritmi modellati sul cervello umano. In questo caso, i dati strutturati non sono necessari perché il sistema è in grado di elaborare anche dati non strutturati. La differenza rispetto al machine learning è che, nel caso del deep learning sono richiesti molti più dati. Inoltre, l'utilizzo del deep learning è particolarmente indicato per compiti complessi, come ad esempio quando non tutti gli aspetti di un fenomeno possono essere categorizzati in anticipo.

Proprio per questo, al fine di catturare in modo più accurato l'andamento di κ_t nel tempo, viene applicata una "rete neurale ricorrente" con un'architettura Long Short-Term Memory (LSTM) e integriamo il modello Lee-Carter per migliorarne la capacità predittiva. In particolare, Nigri et al. (2019) ritengono che l'utilizzo della LSTM consenta di ottenere previsioni di mortalità più coerenti con le dinamiche di mortalità osservate, anche in caso di trend di mortalità non lineari. Questo avviene perché la rete LSTM è strutturata in modo da elaborare lunghe sequenze di dati, formando una memoria in grado di preservare le relazioni significative tra i dati, anche molto distanti nella sequenza. In questo contesto, all'interno delle serie storiche, la rete LSTM permette di prevedere la mortalità futura tenendo conto dell'influenza significativa dell'andamento passato, riproducendolo accuratamente nelle previsioni. Il punto di forza dell'LSTM risiede nella sua capacità di mantenere le informazioni nel tempo, evitando che i segnali più vecchi si disperdano progressivamente durante l'elaborazione.

Pertanto, tale tecnica è focalizzata sulla previsione dell'andamento della mortalità, mentre la stima dei parametri resta quella proposta da Lee e Carter (1992).

3.2.2.1 Modello di Rete Neurale (Neural Network).

Il termine “Neural Network” (NN) deriva da un modello matematico ideato per replicare il funzionamento delle reti neurali biologiche del cervello umano. Questo concetto è stato sviluppato inizialmente da Minsky e Pitts nel 1943 e successivamente approfondito da Wiener nel 1948.

Una rete neurale è generalmente suddivisa in tre strati principali:

- Lo strato di input;
- Lo strato nascosto;
- Lo strato di output.

Ciascuno di questi strati contiene un numero variabile di neuroni. Ogni neurone riceve informazioni “pesate” attraverso connessioni sinaptiche da altri neuroni, e produce un output applicando una funzione di attivazione.

Considerando un singolo neurone H , chiamato “perceptrone” da Rosenblatt (1958), il suo output è definito come:

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

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

Inoltre, il modo in cui le unità sono collegate attraverso le sinapsi determina diversi tipi di reti neurali. Nel modello “classico” di Neural Network (feedforward NN), l’informazione fluisce in una sola direzione, dallo strato di input a quello di output. Al contrario, nelle reti neurali “ricorrenti” (RNN), l’informazione viene elaborata in modo ciclico, utilizzando connessioni sinaptiche aggiuntive che permettono di reinserire il risultato dell’elaborazione come nuovo input.

La seguente figura mostra la tipica rappresentazione di un modello classico di neural network (feedforward NN):

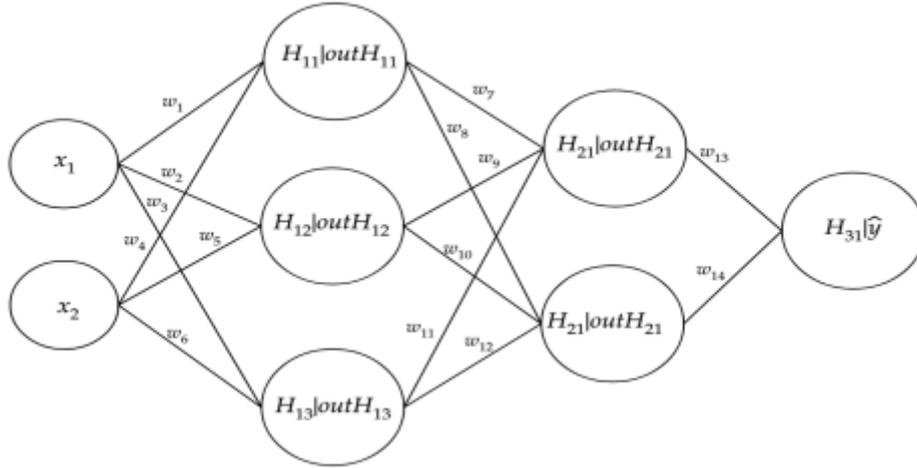


Figura 15; Modello classico di neural network (feedforward NN). Fonte: Nigri et al. (2019): "A Deep Learning Integrated Lee–Carter Model".

Ciò che si osserva è che ogni nodo del grafico rappresenta un neurone. Tutti i neuroni sono collegati tra loro da archi che rappresentano le sinapsi. Nel grafico sono rappresentate le variabili generiche di input, latenti e di output.

L'output ($\mathbf{outH} \in \mathbb{R}^{n_h}$) di un generico strato nascosto è definito come segue:

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

dove $W \in \mathbb{R}^{d \times n_h}$ rappresenta una matrice di peso; $\mathbf{b} \in \mathbb{R}^{n_h}$ è un vettore di distorsioni.

Nello schema MultiLayer Perceptron (MLP), l'output di un livello nascosto diventa l'input per il livello successivo.

Inoltre, indicando con $g \in \mathbb{N}$ il numero di strati nascosti, l'output $\hat{y} \in \mathbb{R}$ si ottiene da:

$$\begin{aligned} \mathbf{outH}_1 &= \phi_1(W_1^T \mathbf{x} + \mathbf{b}_1), \\ \mathbf{outH}_2 &= \phi_2(W_2^T \mathbf{outH}_1 + \mathbf{b}_2), \\ &\dots \\ \hat{y} &= \phi_g(W_g^T \mathbf{outH}_{g-1} + \mathbf{b}_g), \end{aligned}$$

dove: W_1, W_2, \dots, W_g sono matrici di peso; $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_g$ sono vettori di bias; $\phi_1, \phi_2, \dots, \phi_g$ sono funzioni di attivazione non necessariamente diverse tra loro. È fondamentale considerare che le dimensioni delle matrici di peso e dei vettori di bias dipendono dal numero di unità presenti nei livelli nascosti. Quindi, aumentando il numero di livelli nascosti, si incrementa la capacità della rete di rappresentare e comprendere complessità maggiori nei dati di input.

3.2.2.2 Retro-propagazione.

La back-propagation, o anche definita retro-propagazione, è un algoritmo fondamentale nell'addestramento delle reti neurali artificiali. Tale algoritmo è stato introdotto nel 1986 da

Rumelhart, Hinton e Williams. Questo metodo è progettato per ottimizzare i pesi delle connessioni sinaptiche all'interno della rete, riducendo l'errore che intercorre tra le previsioni della rete neurale e i valori target reali. A tal proposito viene introdotta una funzione di perdita che mira a misurare la differenza tra i valori previsti \hat{y} e quelli osservati y :

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

Lo scopo è quello di trovare i valori dei pesi sinaptici che minimizzano la funzione di perdita E .

Tra i molti altri algoritmi, la back-propagation è l'algoritmo più utilizzato per l'addestramento delle reti neurali "classiche". Il processo si svolge in due fasi principali che sono:

1) Fase di **Forward**:

- Durante tale fase i dati di input vengono propagati attraverso i diversi strati della rete neurale, dal livello di input a quello di output, passando per i livelli nascosti. Ogni neurone applica una funzione di attivazione ai segnali in ingresso, che trasforma la somma ponderata degli input in uscita.
- Al termine della funzione di forward, la rete produce una previsione basata sui dati di input, che viene poi confrontata con il valore target per calcolare l'errore.

2) Fase di **Backward**:

- Dopo aver ottenuto la previsione si calcola l'errore confrontando l'output della rete con il valore target reale. Tale errore rappresenta la differenza tra la previsione e la realtà.
- Successivamente, l'errore viene retro-propagato attraverso la rete, dal livello di output verso i livelli nascosti e infine al livello di input. Questo processo utilizza il teorema della catena per calcolare come l'errore si propaga all'indietro, determinando il gradiente dell'errore.
- I gradienti calcolati vengono utilizzati per aggiornare i pesi attraverso la discesa del gradiente, che è un processo descritto dettagliatamente da Haykin (2009). L'idea alla base della discesa del gradiente è simile a "scendere da una collina" fino a raggiungere un minimo globale o locale. Ad ogni aggiornamento, la ricerca si sposta nella direzione opposta al gradiente e la pendenza del gradiente e la velocità di apprendimento determinano l'ampiezza di questo movimento. Questo aggiornamento serve a minimizzare l'errore, migliorando così la precisione della rete.

Il processo di Forward e Backward viene ripetuto per molte epoche attraverso il dataset di addestramento. Con ogni iterazione, la rete apprende e ottimizza i suoi pesi, migliorando la sua capacità di fare previsioni accurate.

La back-propagation ha rivoluzionato il campo delle reti neurali, permettendo di addestrare reti profonde con migliaia di neuroni e strati nascosti, come discusso da Goodfellow et al. (2016) nel libro "Deep Learning". Tale processo iterativo è alla base

dell'apprendimento profondo e costituisce una delle tecniche più importanti per lo sviluppo di modelli di intelligenza artificiale avanzata.

3.2.2.3 Rete Neurale ricorrente con memoria a breve-lungo termine (LSTM).

Seppur le reti neurali semplici rappresentano un potente strumento di analisi (Hornik et al. (1989)), sono inadeguate a gestire efficacemente sequenze temporali di dati. Le connessioni ricorrenti che invece caratterizzano le reti neurali ricorrenti (RNN) permettono di avere un'analisi dinamica dei dati sequenziali. Tuttavia, utilizzando le RNN, ci troviamo di fronte ad un problema principale, ovvero che i gradienti svaniscono, e i pesi diventano molto piccoli senza avere più nessun effetto. Pertanto, la rete neurale perde la sua capacità di apprendere dal passato, diventando inadeguata per l'analisi di lunghe sequenze di dati. Ciò sta a significare che le reti neurali ricorrenti hanno solo una memoria corta.

Per superare tale problema, Hochreiter e Schmidhuber (1997) hanno sviluppato la cosiddetta "Long Short-Term Memory", LSTM. Essa è una rete neurale ricorrente la cui architettura è tale da permettere di considerare le relazioni tra i dati della sequenza, anche se quest'ultima è lunga. In questo modo viene eliminato il problema del gradiente evanescente.

A questo punto, le reti neurali ricorrenti acquisiscono sia memoria lunga che breve, riuscendo a generare delle ottime prestazioni, anche in un contesto di serie storiche. Nel tempo, il lavoro originale riguardante la LSTM, ha subito diversi miglioramenti, come mostrato in Gers et al. (2000).

3.2.2.4 Integrazione della LSTM al modello Lee-Carter.

In questa sezione, integriamo l'architettura LSTM all'interno del modello Lee-Carter tradizionale. L'obiettivo principale è quello di sfruttare i vantaggi offerti dall'LSTM per potenziare la capacità predittiva del modello LC. Per raggiungere tale scopo, è stata ideata una serie di esperimenti finalizzati a valutare l'efficacia dell'LSTM nella previsione dell'evoluzione futura della mortalità e a confrontare le sue prestazioni con quelle ottenute tramite il modello ARIMA.

Pertanto, l'analisi riguarda la previsione di κ_t , mentre i parametri α_x e β_x sono ottenuti secondo la procedura di stima applicata da Lee e Carter (1992).

Diversamente dal modello LC, che impiega un "random walk with drift", adesso si calibra il miglior modello ARIMA(p,d,q) utilizzando l'algoritmo ideato da Hyndman e Khandakar nel 2008. In un primo passaggio, la procedura verifica la stazionarietà delle serie storiche e determina l'ordine di differenziazione (d). Successivamente, vengono selezionati i valori migliori per l'ordine auto-regressivo (p) e per quello della media mobile (q), utilizzando un criterio informativo specifico, come ad esempio l'AIC o il BIC.

Le prestazioni dell'ARIMA vengono poi confrontate con quelle dell'LSTM. Infatti, viene costruito un modello LSTM che approssima la funzione f collegando κ_t ai suoi ritardi temporali, come segue:

$$\mathbf{LSTM}: \kappa_t = f(\kappa_{t-1}, \kappa_{t-2}, \dots, \kappa_{t-j}) + \epsilon_t$$

dove: $j \in \mathbb{N}$ è il numero di ritardi temporali considerati ed ϵ è un termine di errore omoschedastico.

La rete LSTM richiede la suddivisione del set di dati in “set di addestramento” e “set di test”. Il “test di addestramento” viene chiamato anche Supervised Learning, mentre il “test” serve per convalidare il modello. Il dataset relativo al Supervised Learning è costruito come segue:

Output		Input			
κ_t	κ_{t-1}	κ_{t-2}	...	κ_{t-j}	
κ_{t+1}	κ_t	κ_{t-1}	...	κ_{t-j+1}	
κ_{t+2}	κ_{t+1}	κ_t	...	κ_{t-j+2}	
...	
κ_{t+n}	κ_{t+n-1}	κ_{t+n-2}	...	κ_{t-j+n}	

Figura 16: Dataset relativo al Supervised Learning. Fonte: Nigri et al. (2019): “A Deep Learning Integrated Lee–Carter Model”.

Dopo la fase di addestramento, la rete ha appreso la relazione “input-output” e dovrebbe essere in grado di prevedere i valori futuri di κ_t utilizzando solo l’input. I valori previsti di κ_t , al tempo $n + 1, n + 2, \dots, n + j$, vengono eseguiti in modo ricorsivo. In generale, il valore previsto di $\hat{\kappa}_t$ in un tempo generico $n + \tau$ viene calcolato utilizzando i valori di κ_t con $t = (n + \tau - 1, n + \tau - 2, \dots, n + \tau - j)$ come input. Inoltre, è importante notare che, andando avanti nella previsione i valori futuri di κ_t vengono calcolati utilizzando solo i valori previsti di $\hat{\kappa}_t$, e non quelli osservati. Proprio per questi, i dati previsti non sono stati ancora osservati e dunque l’accuratezza del modello diventa impossibile da misurare. L’anno T , che corrisponde all’ultima osservazione nel set di allenamento, è allo stesso tempo il punto di partenza per la previsione. Pertanto, si consideri $T < n$, dove le osservazioni prima di T rappresentano l’insieme di addestramento e il restante $n - T$ forma l’insieme di test.

3.2.2.5 Considerazioni finali.

Nel lavoro presentato da Nigri et al. nel 2019, è stato proposto un modello Lee-Carter potenziato con tecniche di deep learning, basato su una rete neurale ricorrente (RNN) con architettura LSTM, per la previsione dei valori futuri dell’indice κ_t . L’approccio proposto dimostra una precisione della previsione dell’andamento della mortalità significativamente superiore rispetto al modello ARIMA tradizionale. L’LSTM eccelle grazie alla sua architettura che opera attraverso tre fasi che sono: cattura, memorizzazione e replica. Queste caratteristiche permettono all’LSTM di fornire previsioni più accurate dell’andamento decrescente della mortalità nel tempo, superando le prestazioni del miglior modello ARIMA.

Un’altra prerogativa dell’LSTM che è stata evidenziata è che tale strumento di deep learning fornisce ottime informazioni anche in merito all’aspettativa di vita e alla distribuzione dei decessi. Infatti, mentre il modello LC classico e gli altri modelli di mortalità standard tendono a produrre previsioni con un andamento banale e costante nel tempo, l’LSTM riesce a rappresentare meglio la realtà. Ciò permette agli autori di

evidenziare le differenze tra i valori reali e quelli previsti nel tempo. In particolare, l'uso di tale tecnica di deep learning, integrata al modello LC, riduce al minimo la discrepanza tra i valori osservati e quelli previsti dell'aspettativa di vita.

Tutte questi risultati confermano l'utilità dell'integrazione del deep learning nel modello Lee-Carter. Infatti, tale integrazione consente agli studiosi di avere previsioni della mortalità molto più accurate, pur mantenendo i vantaggi del modello LC, come la parsimonia e la robustezza. Come dichiarato da Nigri et al. (2019), quest'ultima caratteristica è particolarmente apprezzata dai professionisti nel settore delle assicurazioni sulla vita, dei piani pensionistici e previdenziali, che devono calcolare e gestire i flussi di cassa futuri, strettamente legati alle dinamiche della longevità. Infatti, poiché le reti neurali sono strumenti flessibili è possibile superare le limitazioni imposte dai modelli ARIMA e ottenere valutazioni più realistiche delle tendenze future.

Occorre però ricordare che, l'approccio appena descritto fornisce una previsione puntuale, mentre l'analisi della variabilità e la costruzione di intervalli di confidenza restano una grande sfida nel campo delle reti neurali. Proprio per questo, alcune estensioni del lavoro presentato da Nigri et al. nel 2019, cercano di estrapolare le potenzialità della LSTM applicata ad altri modelli di mortalità stocastica.

3.2.2.6 Introduzione delle Reti Neurali Convoluzionali (CNN).

Come già ampiamente detto nei paragrafi precedenti, il modello LC ha il vantaggio di essere semplice, ma allo stesso tempo ha lo svantaggio di non essere ottimale per quanto riguarda le prestazioni di previsione. D'altronde, in letteratura c'è un ampio consenso sul fatto che il modello LC e le sue estensioni, funzionano bene per alcuni, ma certamente non per tutti, i dati di mortalità. Proprio per questo, come ampiamente dibattuto negli articoli di Levantesi e Pizzorusso (2019), Levantesi e Nigri (2020) e Nigri et al. (2019), sembra promettente affiancare a tali modelli di mortalità delle tecniche di machine learning. All'interno dell'articolo proposto da Schnurch e Korn (2021), vengono introdotte le reti neurali convoluzionali (neural network convolutional) (CNN).

Fino a qualche anno fa, le CNN non erano ancora state prese in considerazione nella letteratura riguardante la previsione della mortalità. Tali reti neurali sono state introdotte per la prima volta da LeCun et al. (1989). Esse sono progettate per riconoscere e utilizzare le connessioni tra i pixel vicini nei dati di input, come ad esempio le immagini. Questo permette loro di identificare caratteristiche locali, come bordi o forme, e di capire come queste si combinano per formare oggetti più complessi.

Inoltre, nelle CNN i parametri per la convoluzione sono gli stessi, il che rende la rete neurale più efficiente e le stime dei parametri più affidabili. Partendo da questi vantaggi, gli autori Schnurch e Korn utilizzano una rete neurale convoluzionale "bidimensionale" per fare previsioni sulla mortalità. Parallelamente a questo studio, anche lo studio proposto da Perla et al. (2021) utilizza reti neurali per studiare la mortalità. In quel caso però viene utilizzata, oltre ad una rete neurale ricorrente, una rete neurale convoluzionale "unidimensionale". Le CNN unidimensionali fanno sì che il modello esegua le operazioni di convoluzione solo nel tempo e non nella dimensione dell'età. Proprio per questo

Schnurch e Korn ritengono che per catturare le informazioni lungo il piano età-anno sia necessario utilizzare CNN bidimensionali. Ciò è in linea con quanto detto da Meier e Wüthrich (2020), che utilizzano le CNN bidimensionali per rilevare anomalie nei dati di mortalità.

Sempre secondo Schnurch e Korn, nonostante le loro prestazioni predittive tipicamente più elevate, gli studiosi non sempre preferiscono l'utilizzo delle reti neurali perché esse sono difficili da interpretare. C'è da dire che, la capacità di capire perché un modello fa una certa previsione o almeno la convinzione che il modello crei una rappresentazione significativa delle caratteristiche è talvolta considerata più importante delle prestazioni di previsione del modello.

Inoltre, in molte applicazioni non solo è utile fare previsioni accurate sulla mortalità, ma è anche utile o addirittura richiesto quantificare l'incertezza legata a tali previsioni. Per questo, viene considerata una tecnica basata sul bootstrapping. In uno studio pratico si dimostra che tale tecnica fornisce stime affidabili e dettagliate degli intervalli di previsione per la rete neurale convoluzionale. Al contrario, gli intervalli calcolati con il metodo LC tradizionale non sempre riescono a includere i valori reali che si stanno cercando di prevedere.

I principali contributi offerti da Schnurch e Korn (2021) consistono nel:

- Proporre di addestrare una CNN sulla superficie della mortalità per età-periodo.
- Confrontare le previsioni offerte dalla CNN, con quattro parametri di riferimento che sono:
 - Altri due tipi di reti neurali: le "feedforward NN" e le "reti neurali ricorrenti".
 - Il modello ACF e il modello LC.
- Presentare e applicare un approccio bootstrapping per quantificare l'incertezza delle previsioni delle reti neurali.

Una CNN bidimensionale può rilevare relazioni spaziali nei dati che altri modelli potrebbero non essere in grado di sfruttare. Le rappresentazioni intermedie calcolate da questa tipologia di rete neurale si basano sui tassi di mortalità adiacenti l'uno all'altro sia nella dimensione del tempo sia per l'età. In tal modo il modello può incorporare sia la correlazione dei tassi di mortalità lungo la dimensione dell'età, sia gli effetti dell'interazione età-anno di calendario. Tali effetti vengono definiti come "effetti di vicinato" da Wang et al. (2021). Idee simili sono state sviluppate da Denton et al. (2005). Infatti, gli autori hanno scoperto che le correlazioni tra i residui dei modelli di mortalità tendono ad essere elevate per gruppi di età adiacenti. Un semplice esempio di interazioni età-anno sono gli effetti di coorte, che sono spesso presenti nei dati di mortalità (Renshaw e Haberman, 2006).

Allora, è possibile assumere che una caratteristica molto importante delle reti neurali convoluzionali risiede nel fatto che: modelli simili di mortalità, osservati in età diverse o in momenti diversi, dovrebbero portare a output simili.

Lo studio proposto da Schnurch e Korn inizia fissando la lunghezza massima di una finestra temporale che influenza le previsioni: $\tau \in \mathbb{N}$. Seguendo lo studio condotto da Perla et al. (2021) si imposta $\tau = 10$. Successivamente, per tutte le popolazioni $i \in \mathcal{P}$, vengono organizzati i dati sul tasso di mortalità in matrici “età-tempo”:

$$(m_{x,t}^i)_{x \in X_{in}, t=t_i, \dots, t_\tau}, \dots, (m_{x,t}^i)_{x \in X_{in}, t=t_{Y-\tau}, \dots, t_{Y-1}}$$

Tali matrici vengono poi normalizzate elemento per elemento utilizzando il set di dati di addestramento e successivamente utilizzate come input per la CNN. I risultati che la rete viene addestrata a prevedere sono, rispettivamente, i seguenti:

$$(m_{x,t_{\tau+1}}^i)_{x \in X_{out}}, \dots, (m_{x,t_Y}^i)_{x \in X_{out}}$$

In altre parole, la rete viene addestrata sui tassi di mortalità degli ultimi τ anni, per fornire una previsione per l’anno successivo.

In questo modo, si ottengono previsioni per più anni a venire mediante previsioni ricorsive a 1 anno. Ad esempio, si utilizza la previsione dell’anno “1” come input per la previsione dell’anno “2” e così via.

Una CNN è composta da tre tipi principali di strati:

- Gli strati convoluzionali, che generano diverse rappresentazioni degli input.
- Gli strati di pooling, che estraggono i segnali di output più rilevanti dai livelli convoluzionali.
- Gli strati densi, che sono strati feed-forward di base simili a quelli utilizzati nelle FFNN.

In questo modo, i livelli convoluzionali e di pooling di una CNN possono essere visti come potenti estrattori di caratteristiche, i cui output vengono passati a una FFNN, che li utilizza per fare previsioni sul tasso di mortalità.

Sulla falsariga di quanto è stato affermato da Perla et al. (2021), è possibile osservare il modello al quale si fa riferimento. Tale modello si basa sul modello LC ed è il seguente:

$$\log m_{x,t}^i = b_x + \left((W_{x,j})_{j=1, \dots, k}, Z_t^i \right),$$

dove $Z_t^i \in \mathbb{R}^k$ è una funzione non lineare di $(m_{\tilde{x}, \tilde{t}}^i)_{\tilde{x} \in X_{in}, \tilde{t}=t-\tau, \dots, t-1}$.

Tale modello generalizza il modello proposto da Wen et al. (2021):

$$\log m_{x,t}^i = a_x + \sum_{j=1}^k \beta_x^i \kappa_t^{i,j},$$

dove si assume che sia il livello base di mortalità a_x , che gli effetti dell’età $\beta_x^1, \dots, \beta_x^k$, siano identici in tutte le popolazioni considerate. Tuttavia, viene utilizzata una sequenza di strati convoluzionali, di pooling e densi, che possono catturare effetti di interazione complessi e non lineari attraverso la dimensione dell’età e del tempo dei tassi di mortalità.

Quindi, ricapitolando, Schnurch e Korn nel 2021 hanno proposto di effettuare la previsione della mortalità tramite una Rete Neurale Convoluzionale (CNN). Tale tecnica si è dimostrata superiore ai modelli ACF, LC e RNN. Gli autori hanno testato l'affidabilità dei risultati su un periodo di 20 anni, confermando l'efficacia previsionale della CNN. Tuttavia, non esiste un modello che funzioni al meglio per tutte le età, anni o popolazioni. In particolare, un'analisi degli errori specifici per popolazione aiuta a capire se un modello è adatto per una determinata popolazione di interesse. Ad esempio, gli autori notano che la CNN produce previsioni non realistiche per alcune popolazioni con tassi di mortalità molto bassi, motivo per cui viene sconsigliato l'utilizzo di tale tecnica per quelle popolazioni che si trovano ai margini della distribuzione dei dati di addestramento.

Come già ampiamente detto, i modelli di reti neurali (NN) presentano il limite di essere più complessi da comprendere e meno interpretabili rispetto a modelli più tradizionali, come il modello Lee-Carter. Pertanto, Schnurch e Korn, prima di adottare decisioni su un modello "black box", raccomandano di condurre studi preliminari, come ad esempio:

- Verificare la Bontà di adattamento;
- Calcolare le prestazioni di forecasting;
- Valutare l'incertezza della previsione;
- Controllo della robustezza;

Tuttavia, dopo aver effettuato numerosi studi gli autori hanno affermato che il modello CNN può essere valido per molte delle popolazioni considerate. Infatti, si crede che i modelli "black box" possano rappresentare un utile complemento agli approcci classici, come il modello LC o le tavole di vita deterministiche, utilizzate per la modellazione demografica e attuariale. Tutto ciò è supportato dai risultati ottenuti sull'incertezza delle previsioni: infatti la CNN produce intervalli di previsione affidabili e mostra un plausibile aumento dell'incertezza del modello con l'allungamento dell'orizzonte di previsione.

Uno svantaggio di tali tecniche è che i modelli di reti neurali risultano essere molto costosi da un punto di vista computazionale. L'addestramento dei modelli richiede circa 3-4 giorni, con altri 3 giorni necessari per calcolare la varianza del rumore FFNN, se si desidera quantificare l'incertezza della previsione. Tuttavia, poiché la riqualificazione del modello è necessaria solo quando vengono resi disponibili nuovi dati sulla mortalità, che generalmente si aggiornano su base annuale, e dato che ottenere previsioni da un modello già addestrato è un processo molto rapido, questo aspetto non ostacola l'applicabilità pratica di tali modelli.

Infine, Schnurch e Korn affermano che esistono diversi modi per migliorare o estendere i modelli proposti. Infatti, come proposto da Perla et al. (2021), le combinazioni di vari modelli potrebbero essere considerate collegando gli output di un insieme FFNN-CNN. Questa è un'applicazione dello "stacking" (Wolpert, 1992). La media di più modelli diversi, in questo modo, potrebbe anche ridurre il rischio del modello.

Gli autori hanno addestrato sia il modello RNN che quello CNN per effettuare previsioni one-step, e successivamente sono state generate previsioni multistep utilizzando un

approccio ricorsivo. Sebbene questa sia una strategia comune nella previsione delle serie temporali, esistono altre metodologie che meritano di essere esplorate, come discusso da Ben Taieb e Atiya (2016).

4 ANALISI DELLA SOPRAVVIVENZA DEI PENSIONATI.

4.1 Introduzione e obiettivi dell'analisi.

All'interno del quarto capitolo verranno applicate alcune delle tecniche che sono state introdotte nei capitoli precedenti. Si farà riferimento ad un insieme di dati riguardanti i percettori di rendite dell'INPS e grazie a tali dati si cercherà di capire qual è l'andamento nel tempo di due grandezze di sopravvivenza, ovvero l'aspettativa di vita e l'indice e-dagger. Tutto questo verrà fatto considerando solo i pensionati, ovvero gli individui che hanno un'età compresa tra i 60 e i 100 anni. Questo porterà ad una maggiore conoscenza della mortalità e della longevità dei pensionati, fornendo importanti informazioni a coloro che gestiscono i sistemi previdenziali e pensionistici. Infatti, è risaputo come i sistemi pensionistici sono fortemente influenzati dalla mortalità e dalla sopravvivenza degli individui. Ed è qui che entra in gioco il fenomeno del longevity risk. Infatti, tale rischio si materializza quando le aspettative circa la mortalità non si incontrano nell'esperienza della mortalità (Michael Johnson). Come già detto nel capitolo riguardante il longevity risk, tale fenomeno ha un forte impatto sui differenti tipi di polizze assicurative legate alla vita umana e, soprattutto, sul sistema pensionistico.

Inoltre, Emilio Basetti e Carlo A. Favero, nel 2012, hanno sviluppato un articolo intitolato "Measuring the Impact of Longevity Risk on Pension Systems: The Case of Italy", nel quale si afferma che il rischio di longevità per un sistema pensionistico è rappresentato da una longevità superiore rispetto al previsto, che si traduce in un aumento inaspettato della spesa totale per le pensioni. Allora, il lavoro sviluppato dagli autori va a stimare l'impatto del longevity risk sul sistema pensionistico italiano, combinando le previsioni basate su un modello di mortalità di Lee-Carter (1992) con i pagamenti pensionistici previsti per diverse coorti di pensionati. La novità introdotta è rappresentata dall'utilizzo dell'incertezza, generata da un modello di mortalità, per valutare l'impatto del rischio di longevità sulla spesa pensionistica.

L'utilizzo di modelli di mortalità per valutare le politiche previdenziali presenta due vantaggi:

- In primo luogo, i modelli di mortalità generano previsioni per l'evoluzione nel tempo della popolazione in ciascuna coorte; questo è il requisito necessario per valutare l'impatto delle riforme della sicurezza sociale che, non essendo di solito retroattive, colpiscono in maniera diversa, diverse coorti di popolazione.

- In secondo luogo, è molto improbabile che i parametri di un modello di mortalità siano influenzati dalla specifica politica di sicurezza sociale adottata dal governo.

Gli autori si concentrano sul caso italiano, in quanto l'Italia rappresenta uno dei casi più interessanti da studiare. Questo perché, secondo Basetti e Favero (2012), l'economia italiana è caratterizzata da:

- un debito pubblico tra i più elevati al mondo (sia in rapporto al PIL che in termini assoluti);
- un sistema pensionistico tradizionalmente molto generoso;
- un tasso di fertilità tra i più bassi al mondo.

Inoltre, il sistema pensionistico italiano è stato oggetto nel tempo di molte riforme che hanno cambiato i requisiti per accedere alle prestazioni previdenziali.

4.1.1 Il sistema pensionistico in Italia.

Prima di osservare nel dettaglio come è composto il sistema pensionistico in Italia, risulta molto utile definire cos'è la pensione. Una definizione importante di pensione è fornita dall'enciclopedia Treccani, nel quale essa è definita nel seguente modo: "la pensione è una rendita permanente o temporanea che lo Stato o gli istituti di previdenza corrispondono ai lavoratori del settore pubblico o privato, oppure ai loro familiari o ad altri aventi diritto, in relazione a un pregresso periodo di servizio o di attività lavorativa, per lo più in occasione del collocamento a riposo per raggiunti limiti d'età, o anche a seguito di eventi che abbiano determinato la morte o uno stato di grave menomazione fisica o psichica".

Inoltre, la pensione rientra tra le prestazioni garantite dalle cosiddette assicurazioni sociali, ovvero un complesso sistema di garanzie contro i rischi alla persona a carico, parziale o completo, del bilancio statale. Tali assicurazioni risultano essere obbligatorie, finanziate su base pubblica e vanno coprire i seguenti rischi, eventi e bisogni dei lavoratori:

- Pensione di invalidità, vecchiaia o pensioni ai superstiti;
- Infortuni sul lavoro e malattie professionali;
- Malattia;
- Disoccupazione;
- Assegni familiari;
- Maternità;
- Collocamento professionale.

L'obiettivo delle assicurazioni sociali è proteggere assicurativamente i lavoratori e le loro famiglie indipendentemente dall'elevatezza del rischio, mediante l'applicazione di un premio (contributo) uguale per tutti. Il finanziamento avviene sia tramite appositi stanziamenti in bilancio effettuati da parte dello Stato, sia tramite i contributi, versati dal lavoratore e dal datore di lavoro, calcolati in funzione del salario del lavoratore. Così come i contributi, anche le prestazioni generalmente dipendono dal salario.

Inoltre, quando si parla delle assicurazioni sociali è possibile individuare due impostazioni che sono l'impostazione germanica e l'impostazione inglese. Seguendo l'impostazione germanica è necessario l'intervento dello stato per tutelare i lavoratori e prevede la copertura obbligatoria dei vari rischi (malattia, infortunio, invalidità e vecchiaia). Quindi in questo caso c'è una centralità dello Stato. Invece, nell'impostazione inglese lo Stato non deve intervenire in campo economico e sociale a parte la tutela puramente assistenziale. L'impostazione inglese è stata sviluppata prima di quella germanica e il Regno Unito ancora oggi continua a concepire l'intervento dello Stato come assistenza minima ai cittadini. Mentre in Italia, così come in Francia e in Germania, le prestazioni pensionistiche dipendono dal salario del lavoratore con vincoli di solidarietà (tutti i membri della società contribuiscono finanziariamente al Sistema).

Inoltre, in Italia l'assistenza sociale è regolamentata dall'articolo 38 della Costituzione che dice chiaramente che: "ogni cittadino inabile al lavoro e sprovvisto dei mezzi necessari per vivere ha diritto al mantenimento e all'assistenza sociale. I lavoratori hanno diritto che siano preveduti (previsti) ed assicurati mezzi adeguati alle loro esigenze di vita in caso di infortunio, malattia, invalidità e vecchiaia, disoccupazione involontaria. Gli inabili ed i minorati hanno diritto all'educazione e all'avviamento professionale. Ai compiti previsti in questo articolo provvedono organi ed istituti predisposti o integrati dallo Stato".

A questo punto vengono definiti gli "enti previdenziali", ovvero le istituzioni previste dall'art. 38 della Costituzione, che hanno il compito di gestire la previdenza. Come già detto, le attività sono finanziate sia con le imposte specifiche (contributi previdenziali) sia con altri trasferimenti dello Stato. In particolare, gli enti previdenziali gestiscono quella che è comunemente definita "previdenza di primo pilastro", distinta invece dalla previdenza complementare (detta "previdenza di secondo pilastro"), che si attua su base volontaria. La partecipazione a tali enti è obbligatoria per legge, ed è possibile individuarne diverse tipologie:

- Enti di diritto pubblico: INPS-INAIL;
- Casse previdenziali con personalità giuridica di diritto privato.

L'Istituto Nazionale della Previdenza Sociale (INPS) è il principale ente previdenziale del sistema pensionistico pubblico italiano, presso cui debbono essere obbligatoriamente iscritti tutti i lavoratori dipendenti pubblici o privati e la maggior parte dei lavoratori autonomi, che non abbiano una propria cassa previdenziale autonoma. Tornando all'analisi oggetto di studio, si osserva che fanno parte dell'INPS anche il FPLD, il fondo relativo ai coltivatori diretti, coloni e mezzadri e il fondo degli artigiani e commercianti. Inoltre, il decreto-legge del 6 dicembre 2011, n° 201 detto "salva Italia", ha portato l'INPS a gestire circa il 95% dei lavoratori italiani. Questo dimostra quanto in Italia è importante tale ente di diritto pubblico. In particolare, la sua origine risale al 1898 e nel corso degli

anni ha subito numerosi cambiamenti fino ad avere nel 1943, post Seconda guerra mondiale, la denominazione che si conosce tutt'oggi.

Il principale obiettivo dell'INPS è quello di pagare le pensioni e le altre prestazioni previdenziali, tramite un finanziamento che deriva per circa il 70% dai contributi obbligatori, e per il restante 30% dai trasferimenti attuati da parte dello Stato.

Quindi, l'INPS si occupa della liquidazione delle due pensioni oggetto di analisi, ovvero quella di vecchiaia e quella ai superstiti:

- Pensione di vecchiaia: pensione erogata a favore di coloro che terminano l'attività lavorativa al raggiungimento di specifici limiti di età anagrafica, a condizione che vantino un tempo minimo di permanenza nella gestione previdenziale (anzianità contributiva).
- Pensione ai superstiti: trattamento pensionistico, riconosciuto in caso di decesso del pensionato (pensione di reversibilità) o dell'assicurato (pensione indiretta), in favore dei familiari superstiti.

Oltre alle pensioni oggetto di studio, però, l'INPS si occupa anche della liquidazione della:

- Pensione di anzianità: pensione erogata al raggiungimento dei requisiti stabiliti in termini di sola anzianità contributiva.
- Pensione di inabilità: trattamento pensionistico erogato ai lavoratori per i quali viene accertata l'assoluta e permanente impossibilità di svolgere qualsiasi attività lavorativa.
- Pensione di invalidità: trattamento pensionistico erogato ai lavoratori la cui capacità lavorativa è ridotta a meno di 1/3 a causa di infermità fisica o mentale.
- Pensione supplementare di vecchiaia: prestazione economica liquidata, a domanda, al lavoratore che può far valere contribuzione accreditata nell'assicurazione generale obbligatoria per l'invalidità, vecchiaia e superstiti.
- Pensione in convenzione internazionale per il lavoro svolto all'estero: trattamento pensionistico stabilito in base ad apposite convenzioni bilaterali stipulate con lo Stato estero.

Poi, oltre alle pensioni occorre dire che l'INPS si occupa anche di corrispondere le indennità di disoccupazione, malattia, maternità, l'assegno al nucleo familiare e si occupa anche di pagare il TFR al lavoratore nel caso in cui il datore risultasse essere insolvente.

Una volta visto quale ente in Italia si occupa di effettuare il pagamento delle pensioni, è importante analizzare come e perché negli anni si è evoluto il sistema pensionistico in Italia. Per farlo si osserva un report pubblicato dalla COVIP intitolato "L'evoluzione del sistema pensionistico in Italia".

Ciò che si osserva all'interno di tale report è che, in Italia, il sistema pensionistico pubblico è strutturato secondo il criterio della ripartizione, ovvero i contributi versati da lavoratori e aziende vengono utilizzati per pagare le pensioni di coloro che hanno lasciato l'attività lavorativa. Quindi avviene un trasferimento, basato su logiche redistributive e di solidarietà, tra generazioni differenti. È evidente che in un sistema così organizzato, il flusso delle entrate (contributi) deve essere in equilibrio con l'ammontare delle uscite (pensioni). Se ciò non accade e si verifica uno squilibrio, il deficit ricade direttamente sul bilancio dello Stato. In questo caso si parla di "disavanzo".

Nel corso degli ultimi trent'anni il sistema previdenziale italiano è stato interessato da riforme strutturali, finalizzate sia al progressivo controllo della spesa pubblica per pensioni, sia all'istituzione di un sistema di previdenza complementare che si affianchi a quello pubblico. Secondo la COVIP, per comprendere la portata di queste riforme, è importante riassumere le più importanti tappe dell'evoluzione del sistema pensionistico in Italia.

A partire dagli anni '90, sono state avviate riforme strutturali che hanno riguardato il settore pensionistico. In Italia, come già osservato nei capitoli precedenti, in quegli anni c'è stato un progressivo aumento della vita media della popolazione e questo ha portato a pagare la pensione agli individui per un tempo più lungo. Inoltre, sempre negli stessi anni, il rallentamento della crescita economica ha frenato le entrate contributive. Questo ha generato uno squilibrio nel quale le spese erano maggiori rispetto alle entrate contributive. Per far fronte a tale situazione, sono state attuate una serie di riforme orientate a tenere conto delle esigenze di sostenibilità dei conti pubblici. Tali riforme hanno portato ad un innalzamento dei requisiti minimi per ottenere la pensione e, inoltre, l'importo della pensione è stato collegato: all'ammontare dei contributi versati dal lavoratore durante tutta la vita lavorativa; alla crescita del PIL; alla speranza di vita al momento del pensionamento; e all'inflazione.

In ordine cronologico:

- Fino al 1992: il lavoratore iscritto all'INPS percepiva una pensione il cui importo era collegato alla retribuzione percepita negli ultimi anni di lavoro (sistema retributivo). Considerando una rivalutazione media del 2% per ogni anno di contribuzione, per 40 anni di versamenti, veniva erogata una pensione che corrispondeva a circa l'80% della retribuzione percepita nell'ultimo periodo di attività lavorativa (retribuzione pensionabile).
- Riforma Amato (1992): con tale riforma è stata innalzata l'età per la pensione di vecchiaia ed è stata innalzato gradualmente il minimo contributivo da 15 a 20 anni. Inoltre, siccome era ancora in vigore il sistema retributivo, c'è stato un graduale aumento del periodo di calcolo della retribuzione pensionabile (da ultimi 5 anni a ultimi 10 anni per i dipendenti – da 10 a 15 anni per i lavoratori autonomi).

- Riforma Dini (1995): si passa dal sistema retributivo a quello “contributivo”, nel quale l’importo della pensione non dipende più dal valore della retribuzione ma dall’ammontare dei contributi versati dal lavoratore nell’arco della vita lavorativa. Il passaggio dall’uno all’altro regime è avvenuto in maniera graduale, distinguendo i lavoratori in base all’anzianità contributiva. Si sono create tre diverse situazioni:
 - o I lavoratori che a fine 1995 avevano almeno 18 anni di contribuzione mantennero il sistema retributivo.
 - o ai lavoratori con un’anzianità contributiva inferiore ai 18 anni alla stessa data è stato attribuito il regime misto, cioè retributivo fino al 1995 e contributivo per gli anni successivi (metodo contributivo pro-rata).
 - o Metodo contributivo per i lavoratori che hanno iniziato a lavorare a partire dal 01/01/1996.

Inoltre, con tale riforma c’è stato anche un inasprimento graduale dei requisiti per il pensionamento di anzianità.

- Riforma Prodi (1997): sono aumentati i requisiti d’età per l’ottenimento della pensione di anzianità ed è aumentato l’onere contributivo per i lavoratori autonomi.
- Riforma Maroni (2004): tale riforma ha portato ad un aumento dell’età anagrafica per ottenere la pensione di anzianità (60 anni + 35 di anzianità per i dipendenti e 61+35 per gli autonomi) e quella di vecchiaia (65 anni uomini e 60 anni donne). Vengono inoltre fissati i criteri di delega per una riforma che riguarda la previdenza complementare. Elementi cardine di tale delega sono: una migliore equiparazione tra le diverse forme pensionistiche complementari e il conferimento del TFR da parte dei lavoratori dipendenti alla previdenza complementare anche con modalità tacite.
- Riforma Prodi (2007): si introducono le cosiddette “quote” per l’accesso alla pensione di anzianità, determinate dalla somma dell’età e degli anni di contribuzione: nel 2009 la quota da raggiungere era di 95 anni (con almeno 59 anni di età), dal 2011 si passa a quota 96 (con almeno 60 anni di età), mentre dal 2013 si sale a 97 (con almeno 61 anni di età). Inoltre, si rende automatica e triennale la revisione dei coefficienti di calcolo della pensione obbligatoria in funzione della vita media calcolata su dati ISTAT.
- Manovra “Salva Italia” (2011): a partire dal 2012 cambiano i requisiti anagrafici per la pensione di vecchiaia, ferma restando l’anzianità contributiva minima di 20 anni. Per i dipendenti pubblici e privati, per gli autonomi e per le lavoratrici dipendenti del settore pubblico, l’età di pensionamento sale da 66 anni a 66 anni e 7 mesi tra il 2012 e il 2018. Per le lavoratrici dipendenti del settore privato l’età di pensionamento sale da 63 anni e 9 mesi a 66 anni e 7 mesi tra il 2014 e il 2018. Per le lavoratrici autonome (commercianti, artigiane e coltivatrici dirette), l’età pensionabile passa da 60 anni a 66 anni e 7 mesi tra il 2012 e il 2018.

Inoltre, con tale manovra cambiano anche i requisiti contributivi per la pensione anticipata per via dell'aumento della speranza di vita.

Tutte queste novità introdotte dalla manovra "salva Italia" però non valgono per coloro che hanno maturato i requisiti entro il 31 dicembre 2011, per i quali rimangono valide le vecchie norme.

- Legge di bilancio (2017) e successive modificazioni: tale legge introduce nuove misure che consentono ai lavoratori di anticipare il pensionamento rispetto alle normali scadenze di legge, nel limite massimo di 3 anni e 7 mesi. In via sperimentale la legge introduce due nuovi istituti che sono: l'anticipo finanziario a garanzia pensionistica (APE volontario) e l'indennità di accompagnamento alla pensione di vecchiaia (APE sociale). Entrambi riguardano i lavoratori dipendenti, gli autonomi assicurati presso le gestioni speciali (artigiani, commercianti e coltivatori diretti) o presso la gestione separata dell'INPS. Sono esclusi i liberi professionisti iscritti alle casse professionali.

L' APE volontario consiste in un prestito agevolato da parte di banche o assicurazioni che hanno aderito a specifici accordi. Riguarda chi abbia compiuto 63 anni di età, maturi il diritto alla pensione di vecchiaia entro 3 anni e 7 mesi e possa far valere almeno 20 anni di contributi. Il prestito, concesso in percentuale variabile a seconda degli anni dell'anticipo richiesto, dovrà essere restituito ratealmente in 20 anni mediante una trattenuta che l'INPS effettuerà direttamente sulla pensione futura.

L'APE sociale invece consiste in un'indennità a carico dello Stato erogata dall'INPS a lavoratori che la legge ritiene meritevoli di particolare tutela (disoccupati, invalidi, coloro che assistono i disabili e gli addetti a lavori gravosi) che abbiano compiuto 63 anni di età e possono far valere 30 anni di contributi. A differenza dell'APE volontario, che essendo un prestito agevolato prevede il pagamento di interessi, l'APE sociale non comporta alcun costo per il richiedente. Il tetto massimo erogabile è pari a 1500 euro mensili.

- Riforma del 2019 e successive modificazioni: tale riforma ha introdotto la cosiddetta "quota 100", esercitabile dai lavoratori che nel triennio 2019-2021 potevano far valere congiuntamente i requisiti anagrafici e contributivi di 62 anni di età e 38 di contributi, al quale veniva riconosciuta la facoltà di accedere al pensionamento anticipato. Successivamente si è passati alla "quota 102" nel 2022 e poi alla "quota 103" nel 2023.

Inoltre, tramite tale riforma è stato introdotto anche il riscatto agevolato della laurea.

Queste sono state le principali riforme che hanno caratterizzato il sistema pensionistico italiano dagli anni '90 fino al giorno d'oggi.

4.2 Introduzione ai dati.

4.2.1 Descrizione del dataset.

Per effettuare l'analisi si ha a disposizione un dataset, intitolato "Mortalità rendite 1981_2016". Tale dataset racchiude le informazioni relative ai percettori di rendite dall'INPS, ed è suddiviso in due sezioni che racchiudono, in maniera separata, le informazioni relative ai "maschi" e alle "femmine".

Inoltre, per entrambi i sessi vengono osservate varie tipologie di variabili che forniscono importanti informazioni sulla popolazione di riferimento. Tra queste, si osservano sia variabili qualitative sia variabili quantitative, e la differenza è la seguente:

- **Variabili quantitative:** una variabile si dice quantitativa quando le sue determinazioni (dette anche valori o intensità) sono espresse mediante valori numerici. Esempi di variabili quantitative sono il reddito, il numero di figli, il rendimento di un titolo azionario, il voto conseguito all'esame di maturità, ecc. Le variabili quantitative si suddividono a loro volta in "discrete" e "continue". Una variabile quantitativa si dice "discreta" quando può assumere un insieme finito o numerabile di valori, mentre si dice "continua" quando può assumere, almeno in teoria, tutti i valori compresi in un intervallo reale o, in altri termini, può assumere una infinità non numerabile di valori diversi.

Le variabili quantitative "discrete" derivano generalmente da operazioni di conteggio, come ad esempio il numero di componenti delle famiglie, il numero di veicoli circolanti, il numero di dipendenti di un'azienda, ecc.

Le variabili qualitative "continue", invece, derivano spesso da operazioni di misurazione, come ad esempio la temperatura, la statura, il peso, l'altitudine, ecc.

- **Variabili qualitative (o categoriali):** una variabile si dice qualitativa quando le sue determinazioni (dette anche modalità o categorie) sono espresse mediante aggettivi o sostantivi. Esempi comuni di variabili qualitative sono il sesso, lo stato civile, il gruppo sanguigno, il colore degli occhi e dei capelli, il titolo di studio, ecc. Va sottolineato che le modalità di una qualsiasi variabile qualitativa risultano essere sempre incompatibili fra di loro, nel senso che ciascuna di esse non può coesistere con nessuna delle altre.

Le variabili qualitative a loro volta si suddividono ulteriormente in "ordinabili" e "non-ordinabili", in relazione alla possibilità di stabilire o meno un ordinamento naturale delle modalità. Per esempio, il titolo di studio è una variabile qualitativa ordinabile in quanto c'è un ordinamento naturale delle sue modalità, mentre la variabile sesso risulta sparsa.

Detto ciò, si osserva che sono dieci le variabili esplicative che caratterizzano il dataset. Esse sono le seguenti:

- 1) **Anno:** si considerano i dati per gli individui per gli anni dal 1981 al 2016.

- 2) **Età:** il dataset mostra le informazioni per gli individui con età compresa tra 0 e 100 anni.
- 3) **Sesso:** come già detto in precedenza si considerano in maniera separata i maschi, indicati con "M", e le femmine indicate con "F".
- 4) **Gestione:** essa è una variabile qualitativa che va a classificare i diversi tipi di fondi pensionistici o categorie di iscritti per un sistema previdenziale, come l'INPS (Istituto Nazionale della Previdenza Sociale) in Italia. Le variabili sono codificate mediante cinque lettere, dove ciascuna di essa corrisponde a un tipo di gestione o fondo. Ad esempio:
 - "A" = Fondo Pensioni Lavoratori Dipendenti (FPLD) (escluse contabilità separate). Esso rappresenta il fondo pensione per i lavoratori dipendenti non iscritti a contabilità particolari.
 - "B" = Coltivatori diretti, Coloni e Mezzadri. Essi sono lavoratori agricoli autonomi o in forme contrattuali specifiche.
 - "C" = Artigiani e Commercianti: lavoratori autonomi nel settore artigiano e commerciale.
 - "E" = Contabilità separate del Fondo Pensioni Lavoratori Dipendenti. Esso fa riferimento a particolari categorie di lavoratori iscritti al FPLD con contabilità specifica.
 - "F" = gestione per le persone con disabilità civili che ricevono pensioni o sussidi per invalidità.

In sostanza, la variabile "gestione" è utilizzata per identificare a quale fondo o categoria previdenziale appartengono i soggetti. Essa è utile per distinguere le diverse tipologie di beneficiari o lavoratori all'interno di un sistema pensionistico.

- 5) **Esposti al rischio:** essa è una variabile quantitativa che indica la somma delle frazioni d'anno di esposizione al rischio di tutti i percettori di rendita esposti nell'anno di riferimento, relativamente alla categoria considerata.
- 6) **Importi rendite esposti:** anch'essa è una variabile quantitativa che mostra la somma delle frazioni di importo annuo delle rendite relative ai gruppi di percettori esposti al rischio nell'anno statistico di riferimento. L'esposizione è riferita alle somme erogate, considerando, nel caso di posizioni plurime sulla stessa testa, la somma delle rendite annue. L'importo delle rendite/pensioni annue è arrotondato all'unità di euro.
- 7) **Eliminati per decesso (decessi):** variabile quantitativa che indica numero dei percettori di rendite eliminati per decesso nell'anno di riferimento. L'eliminazione deve essere riferita alla testa pensionata, considerando eventuali posizioni plurime sulla stessa testa come unitarie. Il numero è intero.
- 8) **Importi rendite eliminate per decesso:** variabile quantitativa che indica la somma degli importi annui delle rendite relative a gruppi di percettori, uscite per decesso dall'anno di riferimento. L'esposizione è riferita alle somme eliminate per decesso del percettore, considerando, nel caso di posizioni plurime sulla stessa testa, la somma delle pensioni/rendite annue eliminate. L'importo è arrotondato all'unità di euro.

9) **Categoria:** essa è una variabile qualitativa che fa riferimento al tipo di pensionamento o sussidio previdenziale di cui un individuo beneficia. In particolare, la variabile categorizza gli iscritti a seconda del motivo per cui ricevono una pensione. In questo caso è possibile individuare tre categorie che vengono codificate nel modo seguente:

- "V" = vecchiaia: indica le pensioni erogate a chi ha raggiunto l'età pensionabile e si ritira dal lavoro per raggiunti limiti di età (pensione di vecchiaia).
- "I" = invalidità previdenziale: riguarda le pensioni concesse per motivi di invalidità, quando un lavoratore non è più in grado di lavorare a causa di una disabilità riconosciuta.
- "S" = Superstiti da pensionato: fa riferimento alle pensioni destinate ai familiari superstiti di un pensionato deceduto (come vedove, vedovi o orfani). Essa è conosciuta come pensione ai superstiti.

In sintesi, la variabile "categoria" distingue i tipi di pensione in base al motivo per cui viene erogata: pensioni per vecchiaia, invalidità o ai superstiti di pensionato.

10) **Classe d'Importo:** tale variabile rappresenta le diverse fasce di valore della pensione, calcolate in relazione all'importo minimo stabilito. In pratica, classifica le pensioni in base a quante volte il loro importo supera la pensione minima fissata. I dati riportati fanno riferimento alle pensioni dell'anno 2011 e l'importo della pensione percepita viene indicato con "P".

Le classi d'importo sono suddivise come segue:

- Classe "1" = se l'importo della pensione (P) è inferiore a 1,38 volte il minimo. In questa fascia rientrano le pensioni più basse. ($P \leq 1,38 \text{ volte il minimo}$).
- Classe "2" = se l'importo della pensione (P) è compreso tra 1,38 volte il minimo e 2,76 volte il minimo. In altre parole, la pensione è tra 1,38 e 2,76 volte il minimo. ($1,38 \leq P < 2,76 \text{ volte il minimo}$).
- Classe "3" = se l'importo della pensione (P) è compreso tra 2,76 volte il minimo e 4,14 volte il minimo. ($2,76 \leq P < 4,14 \text{ volte il minimo}$)
- Classe "4" = se l'importo della pensione (P) è maggiore o uguale a 4,14 volte il minimo. Questa fascia rappresenta le pensioni più alte. ($P \geq 4,14 \text{ volte il minimo}$)
- Classe "9" = se l'importo della pensione non è disponibile. Tale categoria viene utilizzata solo quando i dati relativi all'importo non sono presenti o non possono essere determinati.

In sostanza questa variabile classifica le pensioni sulla base del loro rapporto rispetto al minimo pensionistico. Le classi 1-4 rappresentano le pensioni di importi crescenti, mentre la classe 9 indica i casi in cui l'importo non è noto o non è stato comunicato.

4.2.2 Pulizia del dataset.

Una volta descritti i dati e le variabili che si hanno a disposizione, vengono applicati dei filtri e vengono mantenute solo specifiche variabili, in maniera tale da mantenere solo

quelli che sono i dati che poi risulteranno essere utili ai fini delle analisi conclusiva. La pulizia del dataset di microdati si articola in varie fasi che sono le stesse sia per i maschi che per le femmine. Le fasi sono le seguenti:

- 1) In primis vengono eliminate le variabili che non vengono utilizzate e che non sono utili ai fini dell'analisi. Le variabili che vengono eliminate sono: **Importi rendite esposti** e **Importi rendite eliminate per decesso**.
- 2) Inoltre, per la variabile **classe d'importo** vengono eliminate le osservazioni che presentano classe "9", ovvero quelle osservazioni per il quale non si hanno informazioni in merito alla classe d'importo della pensione.
- 3) Successivamente, per la variabile "**età**", vengono eliminate le osservazioni che fanno riferimento alle età inferiori ai 60 anni. Questo sta a significare che si prendono in considerazione solo gli individui con un'età compresa tra i 60 e i 100 anni. Così facendo si prendono in considerazione solo gli individui in età pensionabile, escludendo quindi gli individui in età più giovani. Infatti, per le "femmine" si è passati da 87467 osservazioni a 54780 osservazioni. Mentre per i "maschi" si è passati da 84589 osservazioni a 52477 osservazioni.
- 4) A questo punto, per la variabile "**categoria**", vengono eliminati gli invalidi. Questo sta a significare che vengono presi in considerazione solo i pensionati di vecchiaia e i superstiti di pensionato, che sono più caratteristici delle età che si stanno considerando. Allo stesso modo quindi, per la variabile "**gestione**" vengono eliminate le osservazioni che fanno riferimento ad una gestione contrassegnata con la lettera "F", ovvero coloro che sono gli invalidi civili. Così facendo per le femmine si passa da 54789 a 37716 osservazioni. Mentre per i maschi si è passati da 52477 a 33632 osservazioni.
- 5) Successivamente, per la variabile "eliminati per decesso" vanno omesse tutte le osservazioni per il quale i dati non sono disponibili. Allora per le femmine si passa da 37716 a 26925 osservazioni. Mentre per i maschi si passa da 33632 osservazioni a 24307 osservazioni.

A questo punto, conclusa la fase di pulizia del dataset, si procede con il calcolo del tasso di mortalità per le osservazioni rimanenti. Questo passaggio è cruciale e particolarmente delicato, in quanto il tasso di mortalità costituisce la variabile principale con il quale si andranno a calcolare le due grandezze di sopravvivenza oggetto di studio e analisi. Una stima accurata di questo indicatore è fondamentale per garantire la validità dei risultati e la loro interpretazione, poiché eventuali imprecisioni in questa fase potrebbero compromettere l'intera analisi. Di conseguenza, viene prestata particolare attenzione alla corretta esecuzione del calcolo per ottenere risultati affidabili e significativi.

Ora il dataset risulta essere completo e si possono iniziare ad estrapolare le prime informazioni.

4.3 Statistiche descrittive preliminari.

All'interno di questo paragrafo, al fine di approfondire lo studio della mortalità e della longevità vengono effettuate delle statistiche descrittive preliminari che permettono di ottenere le prime informazioni riguardanti i decessi che caratterizzano gli individui della popolazione oggetto di studio.

La prima cosa che viene analizzata è la distribuzione dei decessi:



Figura 17: Distribuzione decessi per sesso, 1981-2016.

Analizzando la distribuzione dei decessi della popolazione di riferimento nell'intervallo temporale compreso tra il 1981 e il 2016, si osserva che la maggioranza dei decessi riguarda le femmine, che rappresentano il 57% del totale. Al contrario, il 43% dei decessi è relativo ai maschi. Questo dato riflette la maggiore longevità femminile, un fenomeno ben documentato nelle statistiche demografiche, che fa sì che le donne costituiscano una quota maggiore della popolazione.

Successivamente, si procede ad analizzare, per entrambi i sessi, le percentuali di decessi riferiti ai percettori di pensione di vecchiaia e ai superstiti. Ciò che si osserva viene riportato nei seguenti grafici:

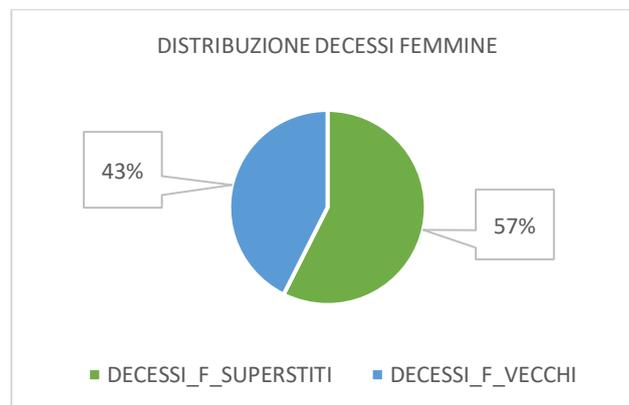


Figura 18: Distribuzione decessi delle femmine per tipologia di pensione percepita.

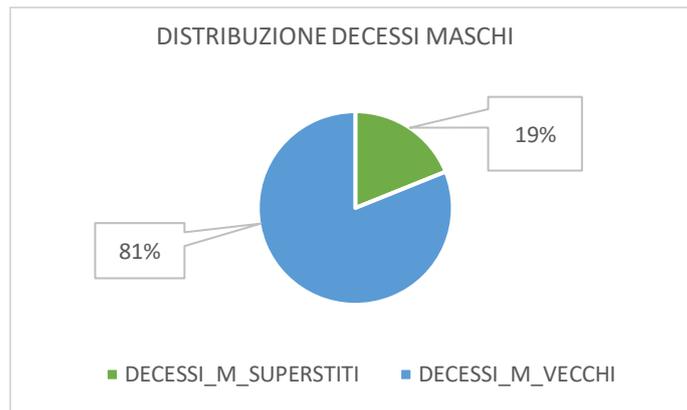


Figura 19: Distribuzione decessi dei maschi per tipologia di pensione percepita

I grafici mostrano la distribuzione dei decessi tra maschi e femmine, suddivisi in base al tipo di pensione percepita al momento del decesso, facendo riferimento all'intervallo temporale 1981-2016. Analizzando i dati relativi agli uomini, si osserva che una netta maggioranza, pari all'81%, era titolare di una pensione di vecchiaia, mentre il rimanente 19% percepiva una pensione di superstite. Questo suggerisce che la maggior parte degli uomini deceduti aveva raggiunto l'età pensionabile e riceveva una pensione direttamente legata alla propria carriera lavorativa.

D'altro canto, per le donne la situazione appare diversa: solo il 43% delle decedute percepiva una pensione di vecchiaia, mentre il 57% riceveva una pensione di superstite, essendo quindi beneficiarie della pensione del coniuge o di un altro familiare. Questa differenza potrebbe riflettere dinamiche storiche e sociali, come un minore coinvolgimento delle donne nel mondo del lavoro in passato o la maggiore aspettativa di vita femminile, che porta le donne a sopravvivere più frequentemente ai loro coniugi e, di conseguenza, a ricevere una pensione di reversibilità.

In sintesi, i dati riferiti al 1981-2016 mostrano un'ampia prevalenza di pensioni di vecchiaia tra gli uomini deceduti, mentre tra le donne decedute prevalgono le pensioni di superstite, evidenziando un'importante differenza di genere nel tipo di pensione percepita al momento del decesso.

Successivamente, si procede con l'analisi dell'andamento dei decessi. In un primo momento, i decessi verranno esaminati considerando l'intera popolazione, senza distinzioni tra i sessi. Successivamente, verrà condotta un'analisi separata per uomini e donne, al fine di valutare eventuali differenze tra i due gruppi.

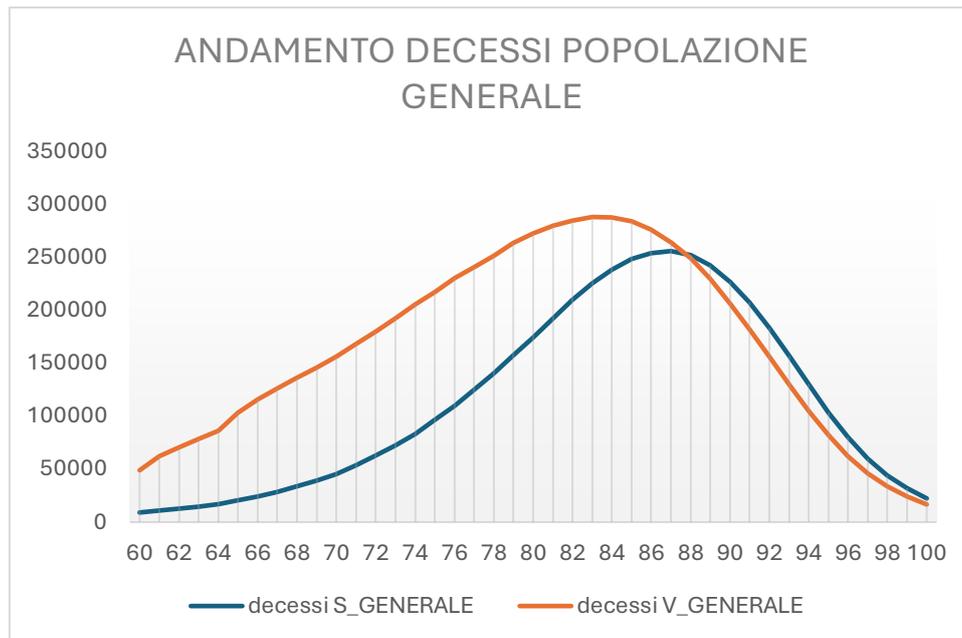


Figura 20: Andamento dei decessi di vecchi e superstiti con età comprese tra i 60 e i 100 anni.

Analizzando la popolazione generale, senza distinzione di genere, e concentrandosi sulla fascia di età compresa tra i 60 e i 100 anni, si osserva una tendenza interessante riguardo ai decessi tra coloro che percepivano una pensione di vecchiaia e i superstiti di pensionato.

Nella fascia di età che va dai 60 agli 88 anni, il numero di decessi tra i pensionati di vecchiaia è significativamente più alto rispetto a quello dei superstiti di pensionato. Questo fenomeno può essere spiegato dal fatto che, in questa fascia di età, la maggioranza degli individui che ricevono una pensione di vecchiaia sono coloro che hanno lavorato e sono andati in pensione, mentre la percentuale di superstiti resta più contenuta.

Tuttavia, dagli 88 anni fino ai 100 anni, la tendenza si inverte. In queste età avanzate, si osserva un maggior numero di decessi tra i superstiti di pensionato rispetto ai pensionati di vecchiaia. Questa inversione può essere dovuta al fatto che, man mano che si sale con l'età, diminuisce il numero di individui che ricevono una pensione di vecchiaia, essendo già deceduti in precedenza, mentre aumentano i superstiti che, in genere, sono coniugi o familiari che continuano a percepire una pensione a seguito della morte del pensionato principale.

Questa dinamica evidenzia non solo il ciclo di vita legato alla ricezione delle pensioni, ma anche l'impatto della longevità, che diventa un fattore chiave nelle età più anziane. I superstiti di pensionato, generalmente più anziani, si ritrovano a costituire la maggioranza dei decessi, indicando come la struttura demografica della popolazione e il ciclo pensionistico si modifichino con il progredire dell'età.

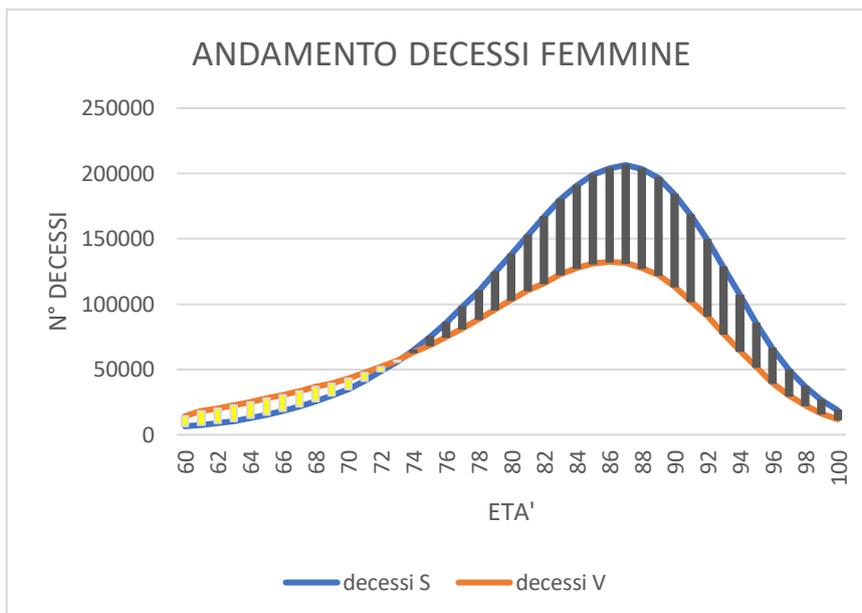


Figura 21: Andamento dei decessi di vecchi e superstiti femmine, con età comprese tra i 60 e i 100 anni.

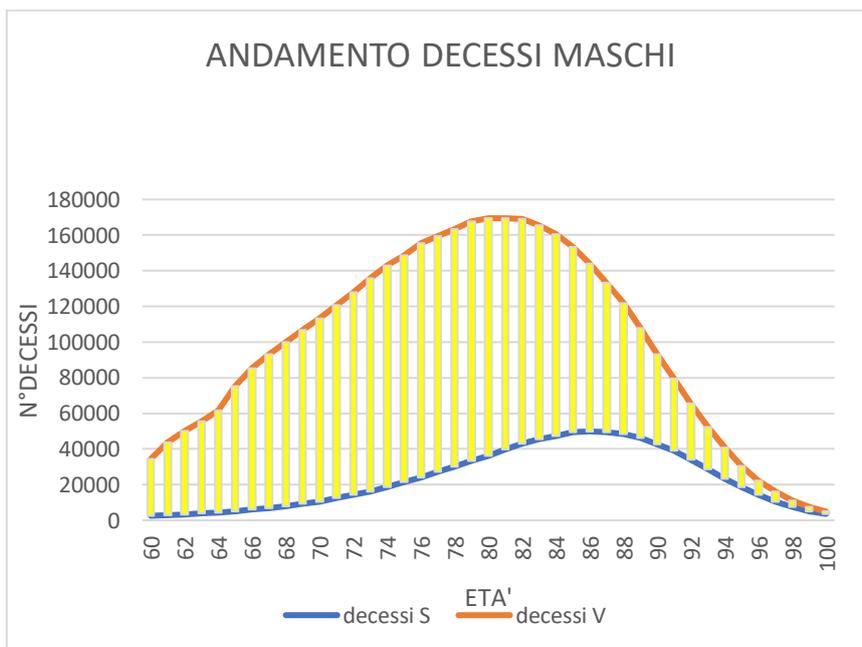


Figura 22: Andamento dei decessi di vecchi e superstiti maschi, con età comprese tra i 60 e i 100 anni.

Quando invece si introduce una distinzione di genere, emergono tendenze differenti per uomini e donne.

Per quanto riguarda gli uomini si osserva che, in tutto l'intervallo di età considerato (dai 60 ai 100 anni), il numero di decessi tra i pensionati di vecchiaia è costantemente superiore rispetto a quello dei superstiti di pensionato. Tale differenza è maggiormente marcata nelle età centrali di tale intervallo. Questo significa che, per gli uomini, la maggior parte dei decessi riguarda coloro che percepiscono una pensione di vecchiaia. Questo potrebbe riflettere il fatto che gli uomini, in generale, hanno una maggiore probabilità di essere i percettori primari della pensione di vecchiaia a causa di una maggiore partecipazione al

mercato del lavoro, mentre il numero di superstiti, tipicamente coniugi o familiari, resta inferiore.

Per le donne, invece si nota un andamento diverso. Tra i 60 e i 73 anni, il numero di decessi tra le pensionate di vecchiaia è maggiore rispetto a quello delle superstiti, rispecchiando in parte la tendenza maschile. Tuttavia, a partire dai 73 anni, la tendenza si inverte: da questa età in poi e fino ai 100 anni, il numero di decessi tra le superstiti di pensionato supera quello delle pensionate di vecchiaia. Questa inversione può essere attribuita a diversi fattori, tra cui il fatto che le donne tendono a vivere più a lungo rispetto agli uomini, e quindi è più probabile che sopravvivano al coniuge, diventando così superstiti beneficiarie di pensione. Inoltre, le donne che ricevono una pensione di vecchiaia tendono a diminuire progressivamente nelle età avanzate, mentre la percentuale di superstiti cresce.

In sintesi, per gli uomini il numero di decessi tra i pensionati di vecchiaia rimane costantemente superiore rispetto a quello dei superstiti lungo tutto l'arco di età considerato. Per le donne, invece, il rapporto cambia nel tempo: inizialmente i decessi tra le pensionate di vecchiaia sono maggiori, ma con il progressivo avanzare dell'età, cresce il numero di decessi tra le superstiti, superando quello delle pensionate di vecchiaia. Questo riflette le differenze demografiche e di aspettativa di vita tra i due sessi e sottolinea l'importanza di considerare il genere nell'analisi delle dinamiche di mortalità, longevità e pensionamento.

A questo punto, i grafici seguenti consentono di esaminare, separatamente per uomini e donne, l'appartenenza dei percettori di pensione a differenti **classi d'importo** e a varie **gestioni previdenziali**. Questa analisi è fondamentale per comprendere come i decessi siano distribuiti in base al livello delle pensioni percepite e alla tipologia di fondo pensionistico a cui gli individui appartenevano.

Attraverso i grafici, si possono identificare eventuali differenze di genere nella distribuzione delle pensioni e nella partecipazione ai diversi fondi previdenziali. Queste informazioni offrono un quadro dettagliato delle condizioni economiche e previdenziali dei pensionati, evidenziando come tali differenze si manifestino tra uomini e donne.

Si inizia a valutare il caso femminile per il quale si osserva che:

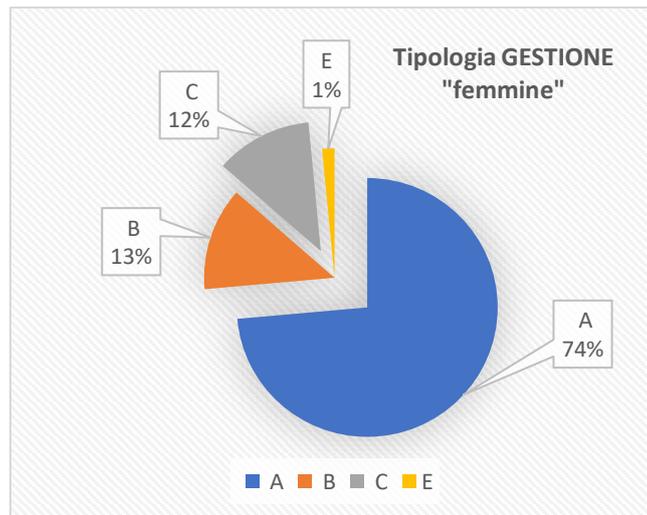


Figura 23: Distribuzione della "gestione" tra le femmine.

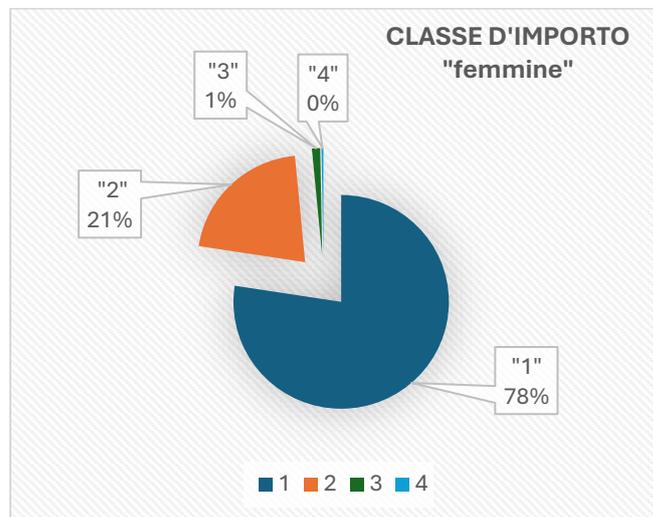


Figura 24: Distribuzione della "classe d'importo" tra le femmine.

Questi dati evidenziano alcune tendenze significative riguardo alla distribuzione delle pensioni tra le donne, sia in termini di gestione previdenziale sia rispetto agli importi percepiti.

- **Analisi delle gestioni previdenziali:**
 Il fatto che il 74% delle donne appartenga alla gestione **A** (Fondo Pensioni Lavoratori Dipendenti) indica che la maggior parte delle donne percepisce una pensione legata al lavoro dipendente. Questo riflette probabilmente una maggiore partecipazione delle donne al mercato del lavoro formale, specie nei settori tradizionalmente coperti da questo fondo. Tuttavia, una quota non trascurabile di donne, pari al 13%, appartiene alla gestione **B** (coltivatori diretti, coloni e mezzadri). Questo suggerisce la presenza di una fetta rilevante di donne che hanno lavorato nel settore agricolo, una categoria storicamente caratterizzata da condizioni di lavoro meno favorevoli rispetto ai dipendenti. Il 12% delle donne che rientra nella gestione **C** (artigiane e commercianti) evidenzia anche un certo livello di partecipazione femminile al lavoro autonomo e imprenditoriale, sebbene resti

minoritaria rispetto al lavoro dipendente. Infine, solo l'1% delle donne appartiene alla gestione E (contabilità separate), un dato che segnala una scarsa presenza femminile in contesti lavorativi con regimi previdenziali speciali.

- **Analisi della classe d'importo:** Guardando ai dati sugli importi pensionistici percepiti, si nota che il 78% delle donne rientra nella classe **1**, con pensioni inferiori a 1,38 volte il minimo, ovvero meno di circa 645 euro al mese (considerando la pensione minima al 2011). Questo dato evidenzia un forte squilibrio economico a sfavore delle donne, che tendono a percepire pensioni molto basse. Ciò potrebbe essere dovuto a diversi fattori, tra cui periodi di carriera più brevi, minori retribuzioni medie, interruzioni di carriera per la cura della famiglia o lavori part-time, tutti fattori che influiscono negativamente sull'accumulo contributivo. Solo il 21% delle donne appartiene alla classe **2**, con pensioni tra 1,38 e 2,76 volte il minimo, che riflette importi leggermente superiori ma ancora modesti (fino a circa 1.290 euro mensili). L'1% delle donne si colloca nella classe **3**, con pensioni comprese tra 2,76 e 4,14 volte il minimo, indicando una piccolissima percentuale di donne che riesce a raggiungere pensioni più consistenti. Infine, nessuna donna appartiene alla classe **4** (pensioni superiori a 4,14 volte il minimo), a conferma del fatto che le pensioni molto elevate sono rarissime tra la popolazione femminile.

Questi dati suggeriscono un forte divario pensionistico di genere, che si traduce in pensioni generalmente basse per le donne, soprattutto nel caso di lavoratrici dipendenti e autonome. Le donne che appartengono alla gestione B o C, legate al lavoro agricolo o autonomo, potrebbero essere particolarmente svantaggiate, riflettendo disparità strutturali nel mondo del lavoro. Anche la scarsa presenza di donne nelle fasce pensionistiche più alte conferma la persistente difficoltà per molte donne di ottenere un'adeguata sicurezza economica in età pensionabile.

Nel complesso, questi numeri richiamano l'attenzione sull'importanza di politiche volte a migliorare l'equità pensionistica, favorendo un maggiore accesso delle donne a carriere stabili e ben remunerate, per correggere le disuguaglianze che si riflettono anche nelle pensioni.

Nel caso degli uomini, come era prevedibile, si osservano dati differenti rispetto a quelli delle donne:

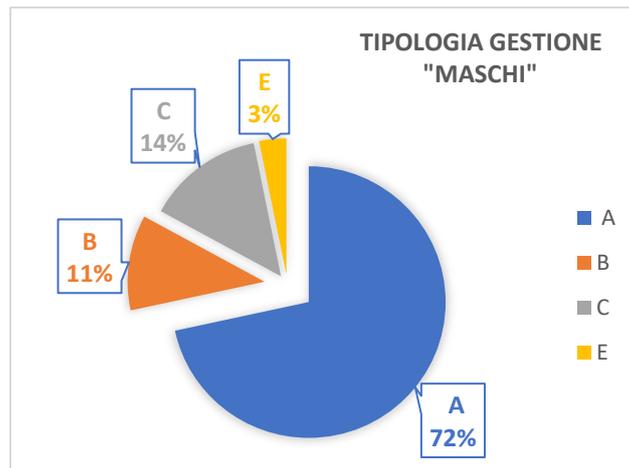


Figura 25: Distribuzione della "gestione" tra i maschi.

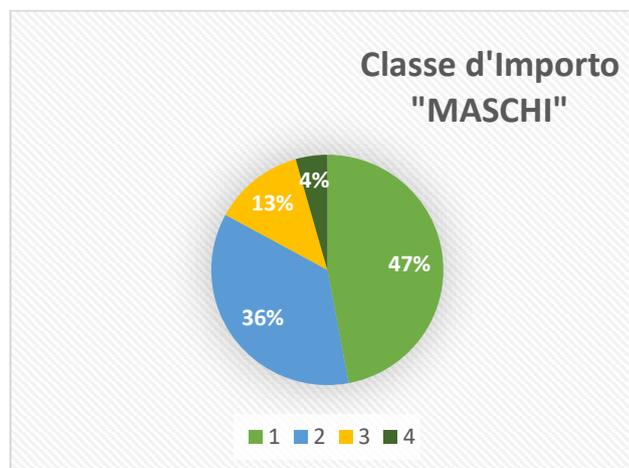


Figura 26: Distribuzione della "classe d'importo" tra i maschi.

Per quanto riguarda gli uomini, il 72% rientra nella gestione **A** (Fondo Pensioni Lavoratori Dipendenti), l'11% nella gestione **B** (Coltivatori diretti, coloni e mezzadri), il 14% nella gestione **C** (Artigiani e Commercianti) e il restante 3% nella gestione **E** (contabilità separate del FPLD). In questo caso la tendenza degli uomini resta approssimativamente simile a quella delle donne, con una netta prevalenza di individui che risultavano essere lavoratori dipendenti.

Invece, in termini di classe d'importo, considerando la pensione minima di 467,43 euro nel 2011, il 47% degli uomini appartiene alla classe **1**, che corrisponde a pensioni inferiori a 645,05 euro. Il 36% degli uomini rientra nella classe **2**, con pensioni comprese tra 1,38 e 2,76 volte il minimo, ovvero tra 645,05 e 1290,10 euro. L'11% degli uomini si colloca nella classe **3**, con pensioni comprese tra 1290,10 e 1935,15 euro, e infine, il 4% degli uomini si trova nella classe **4**, con pensioni pari o superiori a 1935,15 euro.

Questa distribuzione mostra delle differenze significative rispetto a quella delle donne. In particolare, una percentuale maggiore di uomini si colloca in classi di importo più elevate (classi 3 e 4), il che potrebbe riflettere la presenza storica di salari più alti o carriere lavorative più lunghe e stabili per gli uomini, fattori che contribuiscono a una pensione più alta. Questo dato potrebbe essere spiegato dalla disparità di genere nel mercato del

lavoro, dove gli uomini hanno avuto maggiore accesso a impieghi formali e con salari più alti, soprattutto in passato.

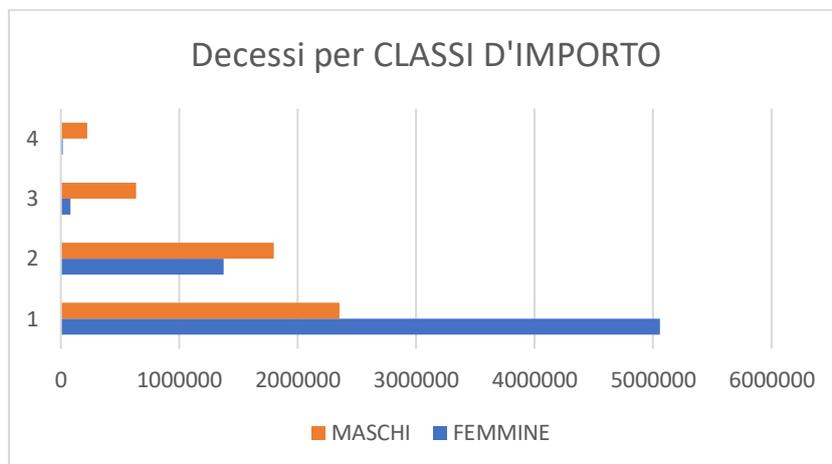


Figura 27: Numero di decessi per classe d'importo, maschi e femmine.

Infatti, effettuando un confronto diretto, si osserva come il numero di femmine è maggiore a quello degli uomini solo per la classe 1, mentre per tutte le altre classi, che denotano una pensione man mano sempre più alta, il numero dei maschi prevale sempre su quello delle femmine.

A questo punto allora si prova a capire che relazione esiste tra la classe d'importo e la gestione.

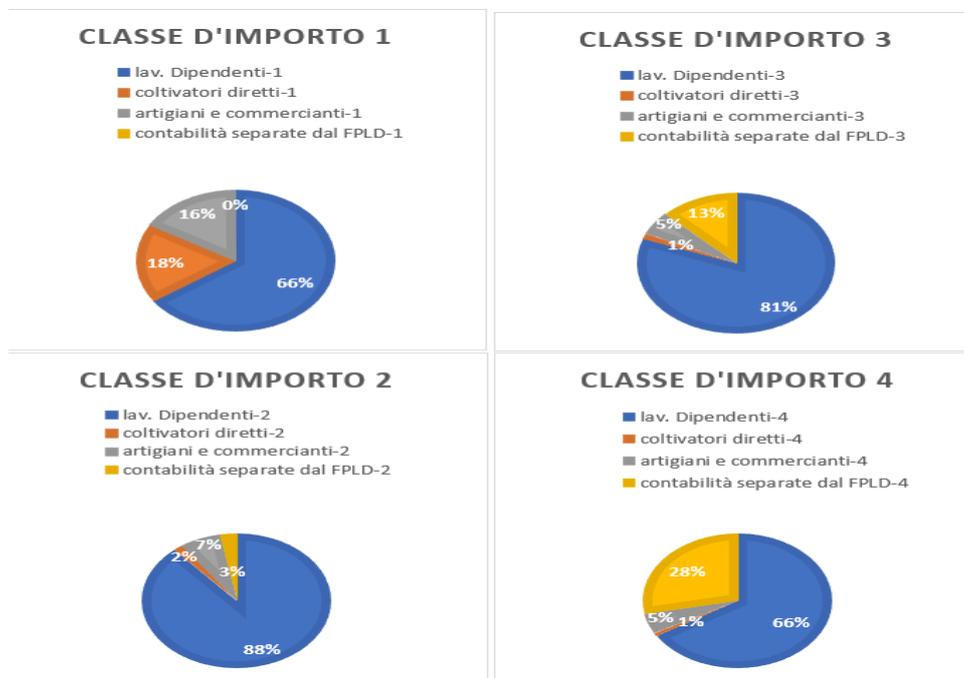


Figura 28: Distribuzione gestione per classi d'importo.

Considerando entrambi i sessi insieme, si osserva una variazione interessante nella distribuzione tra le diverse gestioni pensionistiche in base alla classe d'importo.

Per gli individui appartenenti alla classe d'importo 1 (pensioni inferiori a 645,05 euro), il 66% faceva parte della gestione A (Fondo Pensioni Lavoratori Dipendenti), il 18% della gestione B (Coltivatori diretti, Coloni e Mezzadri), il 16% della gestione C (Artigiani e Commercianti), e nessuno (0%) apparteneva alla gestione E (Contabilità separate del FPLD).

Nella classe d'importo 2 (pensioni comprese tra 645,05 e 1.290,10 euro), l'88% degli individui faceva parte della gestione A, il 2% della gestione B, il 7% della gestione C, e il 3% della gestione E.

Per la classe d'importo 3 (pensioni tra 1.290,10 e 1.935,15 euro), l'81% apparteneva alla gestione A, l'1% alla gestione B, il 5% alla gestione C, e il 13% alla gestione E.

Infine, tra coloro che percepivano una pensione più elevata, ovvero appartenevano alla classe d'importo 4 (pensioni superiori a 1.935,15 euro), il 66% era nella gestione A, l'1% nella gestione B, il 5% nella gestione C, e il 28% nella gestione E.

Tali dati forniscono delle informazioni molto importanti:

- Prevalenza della Gestione A: in tutte le classe d'importo, la maggior parte degli individui apparteneva al Fondo Pensioni Lavoratori Dipendenti. Questo è prevedibile, dato che esso rappresenta il gruppo più ampio della popolazione lavorativa.
- Gestione E e Pensioni Elevate: nelle classi d'importo più alte (3 e 4), si osserva un aumento della percentuale di pensionati appartenenti alla gestione E. Ciò potrebbe indicare che individui con pensioni più elevate tendono a provenire da categorie lavorative che prevedono un trattamento pensionistico diverso e più favorevole.
- Bassa presenza della gestione B e C nelle classi più alte: le gestioni B e C (coltivatori diretti e Artigiani/commercianti) sono sottorappresentate nelle classi d'importo più alte, suggerendo che queste categorie di lavoratori tendono a percepire pensioni più basse rispetto ai lavoratori dipendenti.

Queste distribuzioni, riflettono probabilmente differenze strutturali tra le carriere lavorative e i contributi previdenziali. Infatti si osserva che i lavoratori dipendenti tendono a contribuire di più e quindi a percepire pensioni più alte rispetto ai lavoratori autonomi o agricoli.

Arrivati a questo punto, si osserva qual è stato l'andamento nel tempo, dal 1981 al 2016, del tasso di mortalità. Questo viene visto sia per la popolazione in generale sia per i sessi separatamente.

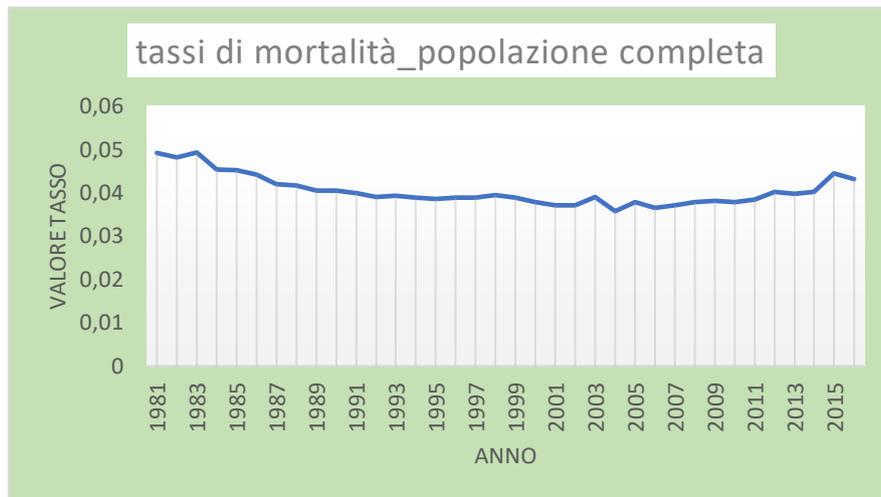


Figura 29: Andamento tasso di mortalità, intera popolazione, 1981-2016.

Analizzando l'andamento del tasso di mortalità dal 1981 al 2016, si osserva un trend interessante. Fino al 2005, il tasso di mortalità mostra una tendenza decrescente, seguita poi da un'inversione di rotta, con un andamento crescente negli anni successivi. La diminuzione iniziale potrebbe riflettere miglioramenti nelle condizioni di vita, nei progressi della medicina e nell'accesso ai servizi sanitari, che hanno portato ad un aumento della longevità e una riduzione della mortalità. Durante questo periodo, si sono verificate anche campagne di prevenzione sanitaria più efficienti e miglioramenti nelle cure per malattie croniche e infettive. Dal 2005, come detto poi è iniziata una crescita e questo potrebbe essere legato a diversi fattori, come l'invecchiamento della popolazione.

Inoltre, emergono picchi significativi nel 1983, 2003 e 2015, mentre il valore più basso viene registrato nel 2004.

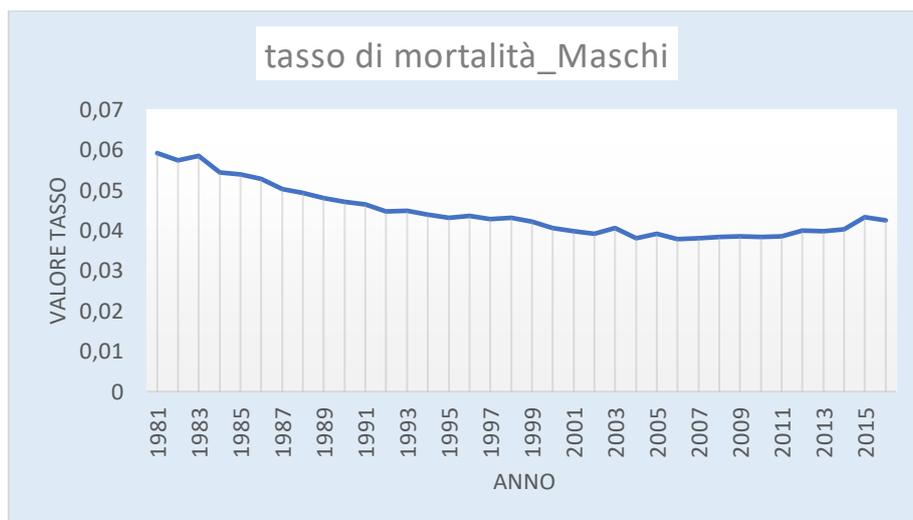


Figura 30: Andamento tasso di mortalità, maschi, 1981-2016.

Analizzando l'andamento del tasso di mortalità solo per gli uomini, si osserva una tendenza decrescente fino al 2008, seguita poi da una crescita negli anni successivi. Questo potrebbe riflettere un miglioramento generale delle condizioni di salute per la popolazione

maschile. Anche un cambiamento negli stili di vita potrebbe aver contribuito a questa riduzione della mortalità.

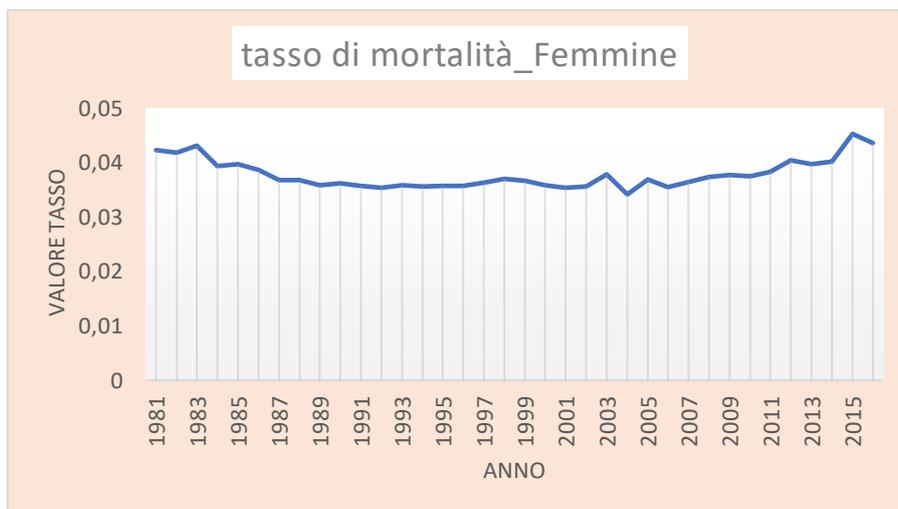


Figura 31: Andamento tasso di mortalità, femmine, 1981-2016.

Al contrario, per le donne, l'andamento risulta molto irregolare, caratterizzato da numerosi alti e bassi, con una forte crescita negli ultimi anni. Tale andamento potrebbe indicare che la mortalità femminile è stata influenzata da una combinazione più complessa di vari fattori come ad esempio il cancro o le malattie cardiovascolari. La forte crescita della mortalità negli ultimi anni potrebbe essere spiegata da un invecchiamento della popolazione femminile, dato che le donne tendono a vivere più a lungo rispetto agli uomini. Questo aumento potrebbe anche riflettere l'effetto di alcuni fattori di rischio, come il cambiamento del ruolo delle donne nella società o l'aumento del fumo e dell'obesità.

Queste differenze nei trend di mortalità tra uomini e donne suggeriscono che i fattori di rischio e i pattern di mortalità siano diversi tra i due sessi, e che l'invecchiamento della popolazione, insieme a fattori ambientali e sociali, stia avendo un impatto crescente su entrambi i sessi.

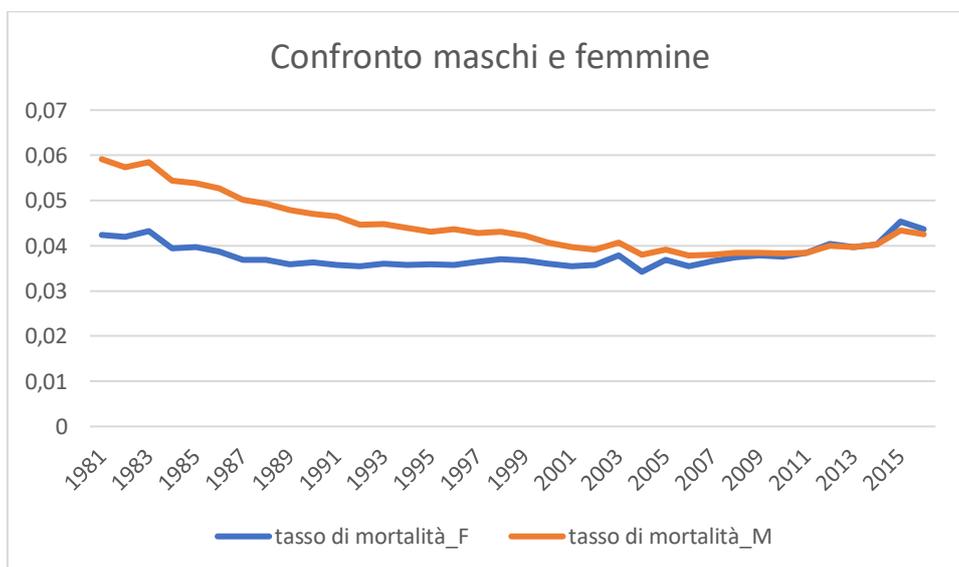


Figura 32: Confronto dell'andamento del tasso di mortalità tra sessi, 1981-2016.

Effettuando un confronto diretto tra i tassi di mortalità di uomini e donne dal 1981 al 2011, si osserva che il tasso di mortalità degli uomini è stato costantemente maggiore rispetto a quello delle donne. Questo può essere attribuito a una maggiore esposizione degli uomini a fattori di rischio come il fumo, l'alcol e lavori fisicamente usuranti o pericolosi. Tuttavia, nel corso di questi anni, la differenza inizialmente ampia tra i due sessi tende progressivamente a ridursi, fino a quasi scomparire nel 2011. Questa graduale riduzione della differenza tra i due sessi potrebbe riflettere la crescente convergenza negli stili di vita tra uomini e donne. Infatti, le donne nel tempo hanno iniziato ad adottare alcuni comportamenti a rischio tipici degli uomini (ad esempio il fumo), il che potrebbe spiegare una parte della riduzione del gap.

Dal 2011 in poi, il tasso di mortalità femminile ha superato quello maschile e ciò potrebbe essere spiegato dall'invecchiamento della popolazione femminile e dai cambiamenti sociali e culturali che ci sono stati tra il XX e XXI secolo, come ad esempio, l'ingresso delle donne nel mondo del lavoro, l'aumento dei livelli di stress o l'esposizione a fattori ambientali.

In generale, la convergenza dei tassi di mortalità tra uomini e donne e il successivo sorpasso da parte delle donne indicano che le dinamiche demografiche e di salute sono cambiate significativamente nel tempo, riflettendo sia l'evoluzione dei fattori di rischio che l'impatto dell'invecchiamento.

4.4 Life Expectancy e Life-span Inequality.

Le statistiche preliminari effettuate nel paragrafo precedente sono utili per individuare le tipologie di individui che sono meglio rappresentate dai dati. Per tali individui si passa al calcolo di due indicatori molto importanti che sono la Life Expectancy, ovvero l'aspettativa di vita (e_x), e la Life-span Inequality, cioè la disuguaglianza nella durata della vita. Per fare ciò, è necessario applicare la function "LifeTable.R" sul software statistico "R". Tale function permette la creazione di una "tavola di vita", uno strumento demografico e statistico utilizzato per analizzare la mortalità di una popolazione.

In questo modo, si è in grado di calcolare la Life Expectancy e la Life-span Inequality che sono due concetti fondamentali in demografia, che offrono una visione diversa della mortalità e della salute in una popolazione.

Già all'interno del primo capitolo è stata introdotta l'aspettativa di vita. Essa è un concetto matematico applicato alla sopravvivenza. È definita come la media degli anni rimanenti da vivere per un individuo o un gruppo. Infatti, va sottolineato che l'aspettativa di vita si riferisce agli anni rimanenti da vivere e non all'età prevista alla morte. In particolare, su questo tema, spesso si fa confusione, perché l'unica informazione ampiamente diffusa sull'aspettativa di vita è quella che viene chiamata "aspettativa di vita alla nascita". Ma, nel caso dell'analisi che si sta descrivendo, si è interessati a calcolare l'aspettativa di vita all'età di 60 anni.

Invece, un concetto nuovo che viene introdotto è quello relativo alla Life-span Variation, anche definita variazione della durata della vita. Tale grandezza è una metrica chiave della mortalità che descrive sia l'incertezza individuale sulla durata della vita, sia l'eterogeneità nella salute della popolazione. Infatti, ciò che si osserva è che l'aspettativa di vita nasconde una notevole eterogeneità nelle traiettorie di mortalità individuali, ovvero le persone non muoiono tutte alla stessa età. (Edwards e Tuljapurkar 2005; Wilmoth e Horiuchi 1999). Questa variazione dell'età alla morte, chiamata *lifespan variation*, è un aspetto cruciale per comprendere la disuguaglianza, perché riflette l'incertezza su quando le persone moriranno.

Storicamente, gli studiosi hanno esaminato le tendenze nella *lifespan variation* nel contesto del declino generale della mortalità a tutte le età. In particolare, con l'aumento dell'aspettativa di vita, l'età alla morte è diventata più prevedibile, il che significa che la *lifespan variation* è diminuita. Questo è accaduto perché la mortalità è diminuita nelle fasce di età più giovani, comprimendo così la maggior parte dei decessi in un intervallo di età più ristretto, principalmente a età avanzate. Tuttavia, non è sempre detto che un aumento dell'aspettativa di vita comporti una riduzione della *lifespan variation*. Questo accade principalmente perché il calo della mortalità nelle fasce di età avanzate continua, mentre la mortalità nelle fasce di età lavorative (adulti giovani e di mezza età) non cala allo stesso ritmo, creando quindi una maggiore variabilità negli anni in cui le persone muoiono. Allora il concetto di *lifespan inequality*, che è anche chiamato variazione della durata della vita, si riferisce alla disparità negli anni di vita vissuti tra gli individui di una popolazione. In una società con bassa *lifespan inequality*, la maggior parte delle persone

muore attorno alla stessa età (ad esempio, quasi tutti vivono fino a 75-85 anni). Al contrario, in una società con alta lifespan inequality, ci sarà una maggiore disparità: alcune persone muoiono giovani, mentre altre raggiungono età molto avanzate.

4.4.1 Analisi Life Expectancy (e_x) e della Life-span Inequality (indice e-dagger e^\dagger).

L'analisi dell'aspettativa di vita a 60 anni (o Life Expectancy) e dell'indice e-dagger offre un'importante chiave interpretativa per comprendere le disuguaglianze demografiche e socioeconomiche all'interno della popolazione anziana pensionata. La Life Expectancy rappresenta il numero medio di anni che un individuo può aspettarsi di vivere dopo aver raggiunto una certa età, in questo caso i 60 anni. L'indice e-dagger, invece, è uno dei principali indicatori della Life-span Inequality, ovvero delle disuguaglianze nella durata della vita. Questo indice misura la perdita media di anni di vita a causa della morte ad una determinata età, riflettendo così la variabilità nella durata della vita all'interno di una popolazione. Un valore elevato dell'e-dagger indica maggiori disuguaglianze nella mortalità tra individui di una determinata coorte.

All'interno di questo paragrafo, si parte dall'analisi delle differenze tra maschi (M) e femmine (F) suddivisi per tipologia di pensione percepita. Si considerano in particolare due categorie di pensioni: la pensione di vecchiaia (V), che riflette una carriera lavorativa completa e regolare, e la pensione di superstite (S), generalmente destinata ai familiari superstiti, in particolare alle vedove e ai vedovi, gruppo spesso caratterizzato da una condizione socioeconomica più vulnerabile. Questo confronto permette di evidenziare come il sesso e la tipologia di pensione influenzino non solo la Life Expectancy, ma anche la Life-span Inequality, espressa attraverso l'indice e-dagger.

Successivamente, l'analisi si estende alla distinzione tra gli individui appartenenti a diverse gestioni previdenziali. A questo scopo, vengono esaminati i tre gruppi principali: gli iscritti al Fondo Pensioni Lavoratori Dipendenti (A), che rappresentano la maggioranza dei lavoratori dipendenti; i coltivatori diretti, coloni e mezzadri (B), storicamente legati al settore agricolo; e gli artigiani e commercianti (C), figure centrali nel settore produttivo e nell'economia autonoma. L'aspettativa di vita e l'indice e-dagger per questi gruppi evidenziano come le diverse condizioni lavorative e previdenziali incidano sia sulla Life Expectancy sia sulle disuguaglianze nella durata della vita.

Infine, l'indagine si concentra sulla combinazione di sesso e classe d'importo della pensione, con l'obiettivo di valutare l'impatto del livello reddituale sulla longevità e sulle disuguaglianze nella sopravvivenza. Le classi d'importo sono suddivise in due fasce: la prima include pensioni fino a 645,05 euro, e la seconda comprende pensioni comprese tra tale cifra e i 1290,10 euro. Attraverso il confronto di questi indicatori, emerge come il livello di reddito e il genere contribuiscano in modo significativo a determinare sia la Life

Expectancy sia la Life-span Inequality all'interno delle varie fasce economiche della popolazione pensionata.

➤ Combinazione sesso-categoria:

la Figura 33 illustra l'andamento dell'aspettativa di vita a 60 anni, dal 1981 al 2016, suddiviso per sesso e per tipologia di pensione percepita. Il grafico permette di evidenziare due aspetti centrali: le differenze di genere nell'aspettativa di vita e l'impatto della tipologia di pensione sulla longevità. Inoltre, consente di osservare il cambiamento di questi indicatori nel corso del tempo, in un periodo storico caratterizzato da significativi miglioramenti delle condizioni sanitarie e di vita che hanno portato a un aumento complessivo dell'aspettativa di vita per tutte le categorie esaminate. Quindi, tale grafico, suddividendo i gruppi per sesso e pensione, facilita il confronto diretto tra le diverse combinazioni e offre una panoramica dettagliata delle disuguaglianze di longevità.

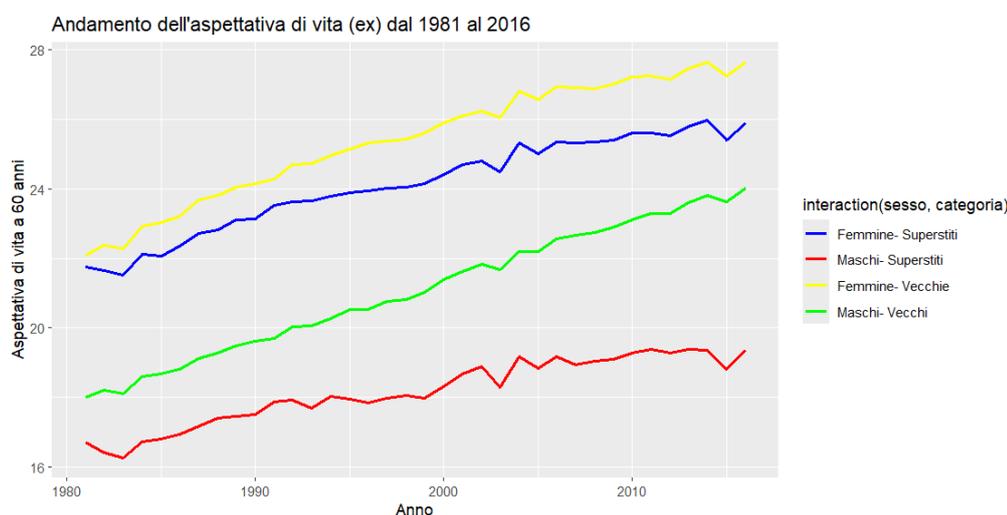


Figura 33: Andamento dell'Aspettativa di vita (ex), dal 1981 al 2016, per la combinazione sesso-categoria.

Le quattro linee presenti nel grafico rappresentano le seguenti combinazioni:

- Linea blu: F-S (femmine, pensione di superstita);
- Linea rossa: M-S (maschi, pensione di superstita);
- Linea gialla: F-V (femmine, pensione di vecchiaia);
- Linea verde: M-V (maschi, pensione di vecchiaia).

In primis, ciò che si osserva è una tendenza generale all'aumento dell'aspettativa di vita. Infatti, tutte le categorie mostrano un progressivo aumento dell'aspettativa di vita a 60 anni nel periodo osservato (1981-2016). Questa tendenza è coerente con il generale miglioramento delle condizioni di vita e sanitarie che hanno caratterizzato gli ultimi decenni in Italia.

In secondo luogo, un altro aspetto fondamentale che emerge da tale grafico è una differenza di genere. Infatti, le donne mostrano un'aspettativa di vita significativamente più alta rispetto agli uomini per entrambe le categorie

pensionistiche. Questo dato riflette una tendenza demografica consolidata, ovvero che le donne tendono a vivere più a lungo degli uomini.

Successivamente, da tale grafico emerge anche una notevole differenza nell'aspettativa di vita tra le diverse categorie di pensione percepita. Infatti, coloro che percepiscono una pensione di vecchiaia, tendono ad avere un'aspettativa di vita più alta rispetto a quelli che ricevono una pensione di superstite. Questo può essere spiegato dal fatto che chi riceve una pensione di vecchiaia ha generalmente una storia lavorativa più stabile e lunga, che potrebbe riflettere una migliore condizione socioeconomica e, di conseguenza, una maggiore longevità. Invece, i percettori di pensione di superstite, tipicamente vedove/i, presentano una minore aspettativa rispetto ai "vecchi", sia per gli uomini che per le donne. Ciò può riflettere condizioni sociali, di salute ed economiche più precarie, in quanto chi riceve una pensione di superstite ha sempre vissuto con e a carico del coniuge prima del decesso.

Un altro aspetto importante da considerare è poi l'evoluzione temporale che si osserva. Infatti, negli anni '80, si osservano gap significativi tra le diverse categorie. Infatti, le donne con pensione di vecchiaia hanno un'aspettativa di vita di circa 4-5 anni più alta rispetto agli uomini con pensione di vecchiaia. Anche per la pensione di superstite, la differenza tra i sessi è marcata. Poi, a partire dagli anni '90 e 2000, tutte le categorie mostrano un incremento dell'aspettativa di vita, ma il gap tra maschi e femmine si riduce lievemente, soprattutto tra i percettori di pensione di vecchiaia. Ciò potrebbe indicare una riduzione delle disuguaglianze tra i sessi in termini di longevità. Poi, nel tempo si osserva anche che la differenza che c'è nell'aspettativa di vita tra i pensionati di vecchiaia e quelli superstiti tende ad aumentare, mostrando un gap che diventa sempre più significativo. Questo avviene sia nel caso degli uomini che nel caso delle donne.

Successivamente, si può osservare anche una particolarità per i maschi percettori di pensione di superstite. Infatti, gli uomini superstiti sembrano mostrare un andamento più irregolare rispetto agli altri gruppi. Questo potrebbe riflettere la vulnerabilità economica e sociale di questo specifico gruppo, tipicamente composto da vedovi.

Quindi, il grafico evidenzia chiaramente le disuguaglianze di genere e le disuguaglianze legate alla tipologia di pensione percepita. Le donne godono di una maggiore aspettativa di vita, indipendentemente dal tipo di pensione, mentre chi percepisce una pensione di vecchiaia ha una longevità maggiore rispetto a chi percepisce una pensione di superstite. Tuttavia, si osserva anche un progressivo miglioramento dell'aspettativa di vita per tutte le categorie, in linea con le tendenze generali dell'invecchiamento della popolazione in Italia.

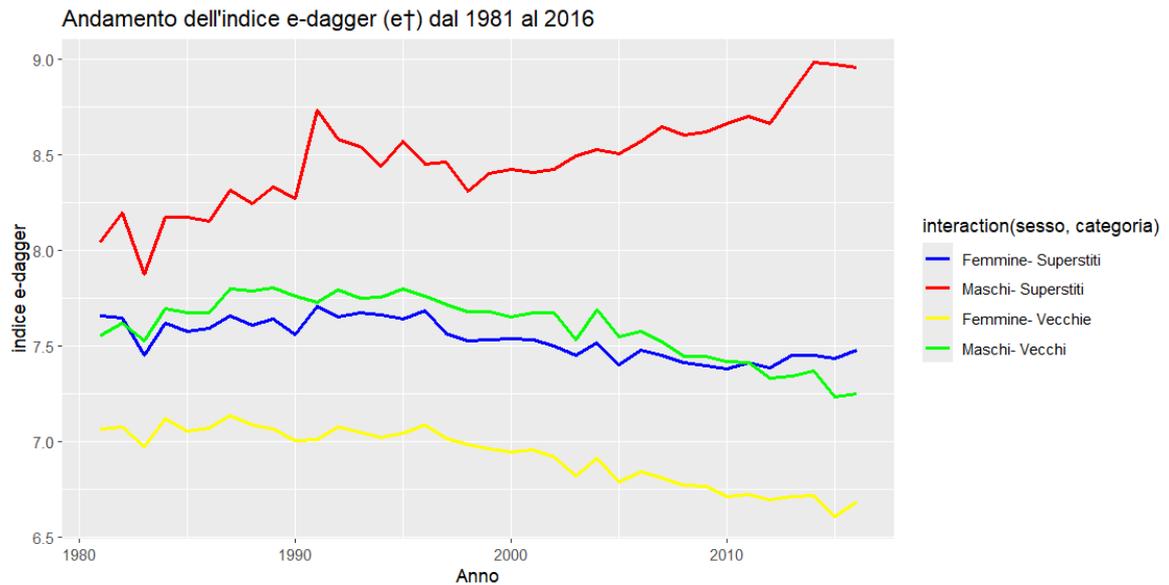


Figura 34: Andamento dell'indice e-dagger dal 1981 al 2016, per la combinazione sesso-categoria.

Successivamente, la Figura 34 rappresenta l'andamento dell'indice e-dagger (e^\dagger) dal 1981 al 2016, suddiviso per sesso e tipologia di pensione percepita. L'indice e-dagger è uno dei principali indicatori della Lifespan Inequality, ovvero delle disuguaglianze nella durata della vita all'interno di una popolazione. Anche in questo caso la suddivisione per sesso e per tipologia di pensione percepita permette di esaminare come fattori demografici e sociali influenzino le disuguaglianze di mortalità nel tempo. Le categorie analizzate sono le stesse che sono state analizzate in precedenza cioè: F-S; M-S; F-V; M-V.

La prima differenza che si osserva è una differenza di genere. Le categorie con la più alta disuguaglianza nella mortalità sono i maschi con pensione di superstite, che presentano l'indice e-dagger più elevato per quasi tutto il periodo, con un valore che supera addirittura 9.0 negli ultimi anni del grafico. Questo suggerisce che tra i maschi "superstiti" esiste una maggiore disuguaglianza nella distribuzione della durata della vita. Il fatto che questa categoria presenti un andamento crescente dal 1995 in poi indica un aumento delle disuguaglianze all'interno di questo gruppo. Al contrario, le femmine con pensione di vecchiaia mostrano i valori più bassi di e-dagger, con un indice che decresce progressivamente da circa 7.2 nel 1981 fino a circa 6.5 nel 2016. Questo suggerisce che tra le donne con pensione di vecchiaia esistono meno disuguaglianze nella distribuzione della mortalità, il che potrebbe riflettere condizioni di vita più omogenee e meno precarie rispetto agli altri gruppi.

Un altro aspetto importante da analizzare è l'evoluzione nel tempo. Infatti, negli anni '80 e nei primi anni '90, le differenze tra le categorie sembravano essere meno marcate. Tuttavia, a partire dagli anni '90, si osserva una divergenza nell'indice e-dagger tra i vari gruppi. Mentre le donne e gli uomini con pensione di vecchiaia

registrano un indice che tende a diminuire o comunque a rimanere relativamente stabile, i maschi con pensione di superstite vedono un aumento significativo delle disuguaglianze di mortalità. Tale aumento è un fenomeno che merita particolare attenzione. Questo potrebbe essere legato a una combinazione di fattori, tra cui il peggioramento delle condizioni economiche o sanitarie per questa fascia di popolazione. Gli uomini che percepiscono una pensione di superstite potrebbero appartenere a gruppi socioeconomici più fragili, con minori opportunità di accesso a cure sanitarie di qualità e condizioni di vita più precarie, il che potrebbe spiegare l'aumento delle disuguaglianze di mortalità.

Un ultimo aspetto da sottolineare riguarda le femmine. Infatti, entrambe le categorie femminili, come detto, mostrano una tendenza alla riduzione delle disuguaglianze di mortalità. Questo potrebbe indicare un miglioramento generale delle condizioni di vita e un aumento dell'eguaglianza in termini di distribuzione della mortalità tra le donne anziane. Anche per le femmine "superstiti" si osserva una tendenza alla riduzione dell'indice e-dagger, pur mantenendosi a livelli più alti rispetto alle donne con pensione di vecchiaia. Ciò può riflettere le difficoltà che le donne vedove devono affrontare.

In conclusione, la Figura 34 sottolinea come la vulnerabilità economica e sociale influisca sulle disuguaglianze di mortalità e sulla variabilità nella durata della vita, con evidenti implicazioni politiche e sociali legate al supporto e alla protezione dei gruppi più fragili.

Effettuando un confronto tra i due grafici si può immediatamente osservare che per i maschi "superstiti" l'aspettativa di vita aumenta nel tempo, però aumenta anche l'indice e-dagger. Questo indica che, pur vivendo più a lungo, all'interno di questo gruppo esiste una crescente disparità nella durata della vita. Ciò potrebbe suggerire che solo alcuni individui all'interno di questo gruppo traggono pieno beneficio dai miglioramenti sociosanitari, mentre altri rimangono indietro, accentuando così le disuguaglianze. L'aumento simultaneo dell'aspettativa di vita e dell'e-dagger suggerisce che l'allungamento della vita non è distribuito equamente, con un sottoinsieme di individui che vive molto più a lungo degli altri. Invece, per le altre combinazioni di individui (M-V; F-V; F-S) l'aumento dell'aspettativa di vita è accompagnato da una riduzione dell'indice e-dagger. Questo è particolarmente marcato nelle donne che percepiscono una pensione di vecchiaia e ciò può essere dovuto ad un miglioramento delle condizioni di vita e di salute, con un impatto favorevole soprattutto sulle donne.

Questo confronto sottolinea l'importanza di analizzare non solo quanto vivono a lungo le persone, ma anche come la durata della vita si distribuisce tra i vari individui all'interno di ogni gruppo, evidenziando le disuguaglianze che possono sorgere nonostante i miglioramenti complessivi nella longevità.

➤ Combinazione sesso-gestione:

La Figura 35 mostra l'andamento dell'aspettativa di vita a 60 anni dal 1981 al 2016, suddiviso per sesso e per gestione previdenziale di appartenenza. I gruppi considerati sono tre:

- A: Fondo Pensioni Lavoratori Dipendenti (FPLD);
- B: Coltivatori diretti, coloni e mezzadri;
- C: Artigiani e commercianti.

Questi tre gruppi sono ulteriormente suddivisi tra maschi (M) e femmine (F), per un totale di sei combinazioni: A-F, B-F, C-F, A-M, B-M, C-M. La seguente figura intende esplorare come la longevità attesa a 60 anni sia influenzata sia dal genere che dalla tipologia di gestione previdenziale.

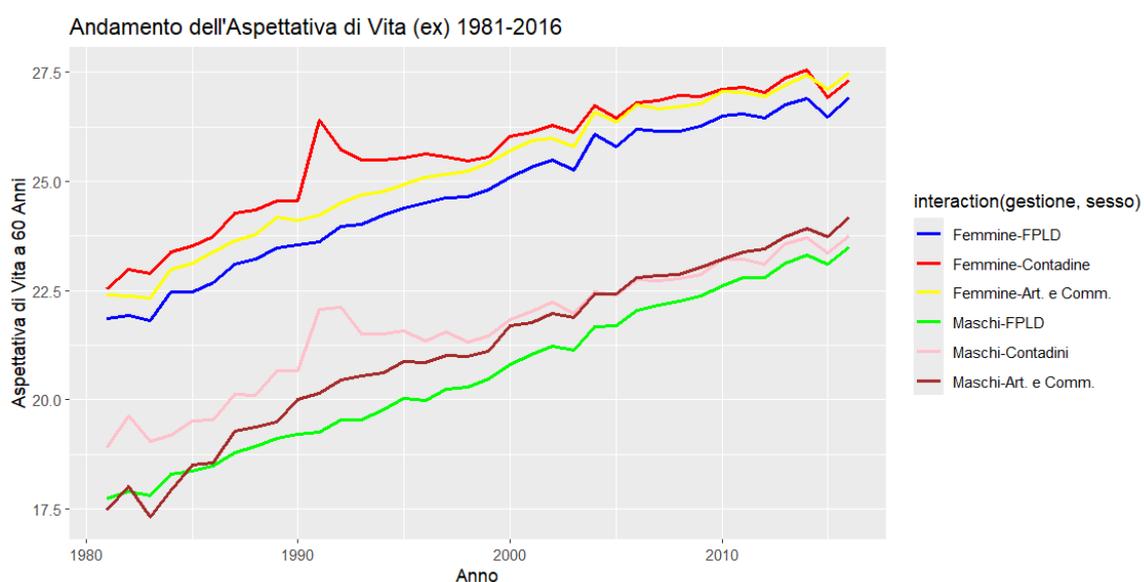


Figura 35: Andamento dell'aspettativa di vita tra il 1981 e il 2016 per la combinazione sesso-gestione.

In primis, ciò che si evince è che persiste una tendenza generale all'aumento. Infatti, tutti i gruppi mostrano una tendenza positiva, con un aumento costante dell'aspettativa di vita a 60 anni tra il 1981 e il 2016. Questo andamento è coerente con i progressi generali nelle condizioni sanitarie e di vita.

Anche in questo caso, così come già analizzato in precedenza, le femmine presentano costantemente un'aspettativa di vita più elevata rispetto ai maschi in tutte le gestioni previdenziali.

Ma ciò che è interessante osservare adesso sono le differenze tra gestioni. In particolare, i lavoratori dipendenti (A) sia maschi che femmine tendono ad avere un'aspettativa di vita minore rispetto agli altri gruppi. Invece inizialmente, i coltivatori diretti (B) avevano valori significativamente più elevati rispetto a quelli degli altri gruppi, con picchi di crescita importanti registrati agli inizi degli anni '90. Questo valeva sia per le donne che per gli uomini. Però, per gli uomini agli inizi degli anni 2000 c'è stato un cambiamento che ha portato gli artigiani e commercianti

(C) ad avere un'aspettativa di vita più elevata rispetto agli altri gruppi. Questo cambiamento è avvenuto anche per le donne, ma circa un decennio dopo. Invece la Figura 36 rappresenta l'andamento dell'indice e-dagger dal 1981 al 2016, suddiviso sempre per sesso e gestione previdenziale. L'indice e-dagger misura la disuguaglianza nella distribuzione della mortalità, e come già detto valori più elevati mostrano maggiori disuguaglianze. Tale Figura si propone di mostrare come le disuguaglianze siano cambiate nel tempo all'interno delle diverse categorie di gestione e sesso.

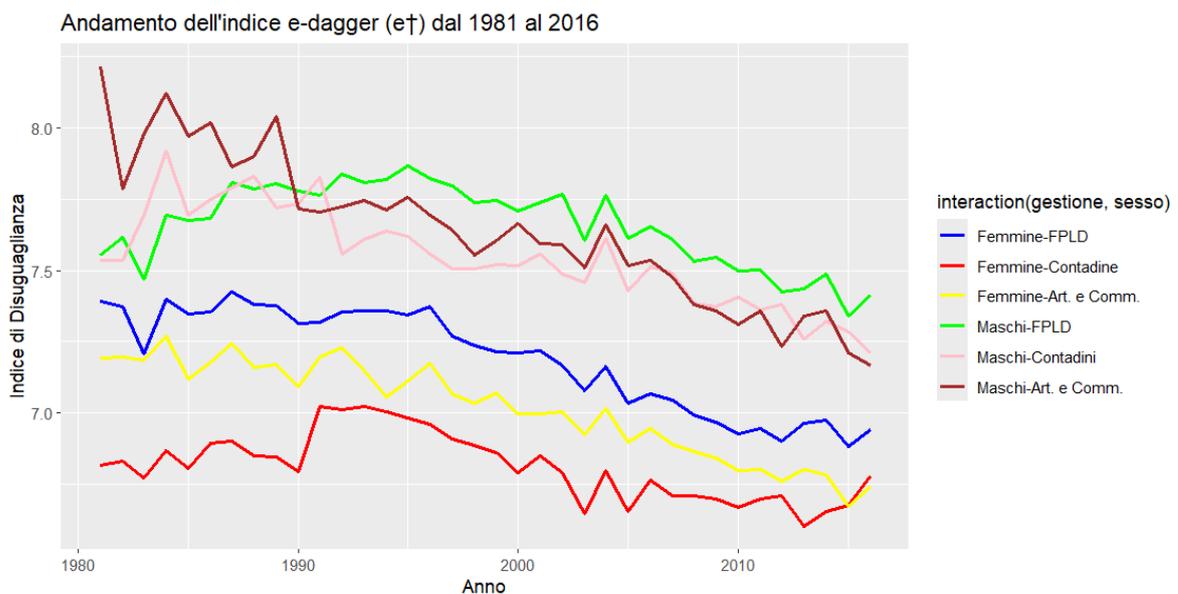


Figura 36: Andamento dell'indice e-dagger tra il 1981 e il 2016 per la combinazione sesso-gestione.

All'interno della Figura 36 si evince che c'è una riduzione generale delle disuguaglianze. Infatti, per quasi tutte le combinazioni, l'indice e-dagger mostra una tendenza alla diminuzione nel corso del tempo, segnalando un miglioramento nella distribuzione della mortalità. Questo indica che la durata della vita è diventata più uniforme all'interno di ogni gruppo.

Anche qui, si evince che c'è una differenza di genere nel quale le donne in tutti i gruppi tendono ad avere un indice e-dagger più basso rispetto agli uomini, il che implica una maggiore equità nella distribuzione della durata della vita tra le donne rispetto agli uomini.

Inoltre, emergono importanti informazioni anche per le diverse tipologia di gestioni previdenziali. Infatti, gli uomini artigiani e commercianti mostrano i valori più alti di e-dagger, specialmente negli anni '80 e all'inizio degli anni '90, indicando che le disuguaglianze nella durata della vita erano molto accentuate in questo gruppo. Tuttavia, c'è una tendenza alla riduzione verso la fine del periodo considerato, infatti, tale gruppo nel 2016 è quello che ha registrato, per gli uomini, un valore più basso, invertendo totalmente la tendenza. Invece, per quanto riguarda le donne

artigiane e commercianti, pur avendo un valore di e-dagger più basso rispetto agli uomini dello stesso gruppo, mostrano anche loro un miglioramento delle disuguaglianze verso la fine del periodo. I dipendenti, sia uomini che donne, mostrano livelli più elevati rispetto agli altri gruppi. Per le donne questo si verifica su tutto il periodo considerato, mentre per gli uomini solo a partire dalla metà degli anni '90.

Infine, effettuando un confronto tra l'andamento dell'aspettativa di vita con quello dell'indice e-dagger, emergono alcune relazioni importanti tra la longevità e la disuguaglianza nella durata della vita. In generale, si osserva che l'aumento dell'aspettativa di vita è accompagnato da una riduzione dell'indice e-dagger. Questo suggerisce che, oltre a vivere più a lungo, le persone in questo gruppo sperimentano anche una distribuzione più equa della longevità, con meno variazioni tra gli individui. Questo è particolarmente evidente per le donne. Inoltre, si osserva anche una convergenza tra uomini e donne verso la fine del periodo. Quindi in conclusione, dall'analisi di questi due grafici, si può affermare che le condizioni socioeconomiche e previdenziali influenzano sia la longevità che l'equità nella sua distribuzione.

➤ **Combinazione sesso- classe d'importo:**

La Figura 37 mostra l'aspettativa di vita a 60 anni e permette una valutazione dell'evoluzione temporale di questa grandezza tra il 1981 e il 2016 in base alla combinazione sesso-classe d'importo. Le classi d'importo sono suddivise in:

- Classe 1: pensioni fino a 645,05 euro;
- Classe 2: pensioni comprese tra 645,05 e 1290,10 euro.

Il grafico analizza come l'aspettativa di vita vari tra maschi e femmine in queste due fasce di reddito, evidenziando le differenze dovute non solo al genere ma anche all'importo della pensione.

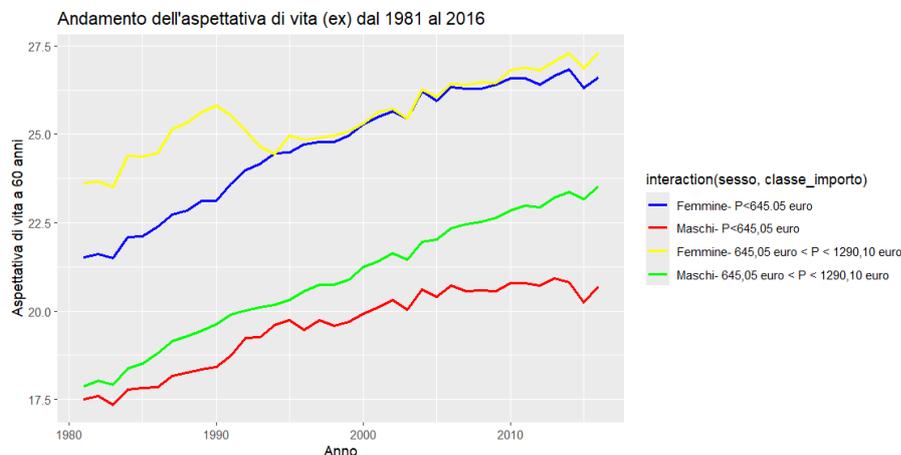


Figura 37: Andamento dell'aspettativa di vita tra il 1981 e il 2016 per la combinazione sesso- classe d'importo.

Anche in questo caso, così come per le combinazioni precedenti, si osserva una differenza di genere nel quale le donne mostrano un'aspettativa di vita più elevata rispetto ai maschi in entrambe le classi d'importo.

In particolare, gli individui della classe 2, sia maschi che femmine, mostrano un'aspettativa di vita più alta rispetto agli individui appartenenti alla classe 1, il che suggerisce una relazione positiva tra il reddito pensionistico e la longevità. Chi percepisce pensioni più alte tende a vivere più a lungo, probabilmente a causa di migliori condizioni di vita e un accesso più agevole a cure sanitarie di qualità. In particolare, le femmine appartenenti alla classe 2 hanno la maggiore aspettativa di vita, con valori che superano i 27 anni dopo i 60 anni. Al contrario, i maschi della classe 1 registrano i valori più bassi, con un'aspettativa di vita inferiore ai 21 anni nel 2016.

Le tendenze temporali sono ben osservabili. L'aspettativa di vita è aumentata per tutti i gruppi negli anni, con una tendenza generalmente costante. Tuttavia, si osserva una leggera convergenza tra i sessi a partire dagli anni 2000, soprattutto per i percettori di pensioni più alte (classe 2).

Successivamente, la Figura 38 mostra l'andamento dell'indice e-dagger nel periodo 1981-2016, considerando la combinazione sesso-classe d'importo. Tale figura ci permette di vedere come tale indicatore della disuguaglianza nella durata della vita possa cambiare in funzione del livello della pensione.

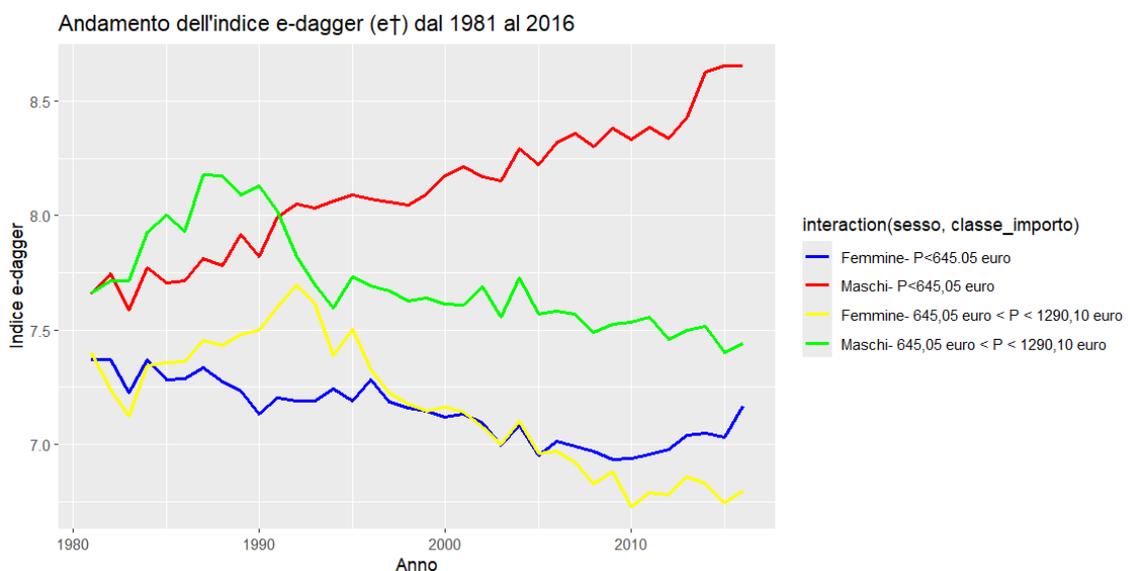


Figura 38: Andamento dell'indice e-dagger tra il 1981 e il 2016 per la combinazione sesso-classe d'importo.

Per quanto riguarda gli uomini si osserva che l'indice e-dagger per i maschi della classe 1 è quello più elevato e mostra un trend in crescita, soprattutto negli anni più recenti. Questo indica che tra i maschi con pensioni basse, le disuguaglianze nella durata della vita stanno aumentando: alcune persone vivono significativamente più

a lungo di altre, accentuando la disparità. Al contrario, i maschi della classe 2, pur avendo un indice e-dagger più basso rispetto a quelli della classe 1, mostrano una tendenza più stabile. Le femmine, in entrambe le classi, presentano valori generalmente più bassi rispetto ai maschi. Ciò suggerisce che, oltre ad avere un'aspettativa di vita più elevata, le donne tendono ad avere una distribuzione della durata della vita più uniforme rispetto agli uomini. Tra le donne, quelle appartenenti alla classe 1, presentano un livello di disuguaglianza maggiore rispetto a quelle della classe 2 verso la fine del periodo considerato.

In conclusione, effettuando un confronto tra i due grafici, si osserva come i percettori di pensioni più elevate, sia maschi che femmine, godono non solo di una maggiore aspettativa di vita, ma anche di disuguaglianze più contenute. La classe 2 mostra valori più bassi di e-dagger, indicando che la longevità è distribuita in modo più uniforme tra gli individui con pensioni più alte. Questo trend suggerisce che una maggiore sicurezza economica contribuisce a migliorare sia la longevità che l'equità nella distribuzione della durata della vita, probabilmente grazie a migliori condizioni di vita. Al contrario, tra i maschi della classe 1, l'aumento dell'aspettativa di vita è accompagnato da una crescita significativa delle disuguaglianze, come indicato dall'indice e-dagger in forte aumento dopo il 2000. Ciò significa che, sebbene alcuni individui in questo gruppo stiano vivendo più a lungo, altri continuano a essere penalizzati, accentuando le disuguaglianze interne. Questo potrebbe riflettere condizioni economiche precarie che non consentono a tutti di beneficiare degli stessi miglioramenti sanitari e sociali.

In conclusione, le precedenti analisi dell'aspettativa di vita a 60 anni e dell'indice e-dagger, in relazione alle diverse combinazioni di variabili come sesso, categoria pensionistica, gestione previdenziale e classe d'importo, hanno messo in luce dinamiche complesse e interconnesse che evidenziano significative disuguaglianze di longevità all'interno della popolazione di percettori di rendite dall'INPS.

Le femmine, in quasi tutte le combinazioni analizzate, mostrano una aspettativa di vita più alta rispetto ai maschi, confermando una tendenza consolidata. Tuttavia, nonostante le donne vivano più a lungo, anche il loro livello di disuguaglianza nella durata della vita, misurato dall'indice e-dagger, è generalmente più basso rispetto agli uomini. Questo significa che tra le donne, la longevità è distribuita in modo più uniforme, suggerendo che le donne tendono a beneficiare più equamente dei miglioramenti in salute, economici e di condizioni di vita.

Inoltre, la categoria e la gestione previdenziale di appartenenza influiscono significativamente sia sulla longevità che sulla distribuzione delle disuguaglianze, i percettori di pensione di vecchiaia tendono ad avere un'aspettativa di vita più alta e una minore disuguaglianza rispetto ai percettori di pensione di superstita. Allo stesso modo, i

lavoratori appartenenti alla classe B (coltivatori diretti, coloni e mezzadri) mostrano migliori performance in termini di sopravvivenza rispetto a gruppi più vulnerabili.

La classe d'importo gioca un ruolo cruciale nel determinare le due grandezze di sopravvivenza analizzate. I percettori di pensioni più elevate (classe 2) godono di una vita più lunga e di una maggiore equità nella distribuzione della mortalità, con l'indice e-dagger che tende a essere più basso e stabile nel tempo. Al contrario, tra i percettori di pensioni più basse (classe 1), specialmente per i maschi, si osserva un'aspettativa di vita più corta e un aumento delle disuguaglianze, con un indice e-dagger in crescita, indicando una distribuzione meno equa della durata della vita.

I maschi risultano essere il gruppo più vulnerabile in termini di Lifespan Inequality. Nonostante un aumento dell'aspettativa di vita negli ultimi decenni, l'indice e-dagger per gli uomini, soprattutto per quelli con pensioni più basse e in categorie meno protette (dipendenti e percettori di pensione di superstita), continua a crescere. Questo suggerisce che gli uomini, specialmente nelle fasce economicamente più deboli, stanno vivendo una crescente disparità, con individui che traggono più vantaggi rispetto ad altri dai miglioramenti generali in salute e condizioni di vita.

L'analisi complessiva ha quindi evidenziato come genere, tipologia di pensione, gestione previdenziale e reddito pensionistico influenzino in modo significativo non solo quanto a lungo le persone vivono, ma anche quanto equamente la longevità è distribuita all'interno di ogni gruppo. Mentre l'aumento dell'aspettativa di vita è un dato generalmente positivo, è chiaro che non tutti i gruppi ne beneficiano in modo equo. Le disparità economiche e le condizioni socioprofessionali giocano un ruolo cruciale nel determinare la variabilità nella durata della vita, con i gruppi più vulnerabili, come i maschi con pensioni basse che restano più esposti ad eventuali disuguaglianze.

Questo studio ha come obiettivo quello di suggerire l'importanza di politiche pubbliche mirate a ridurre le disuguaglianze nella sopravvivenza, promuovendo una maggiore equità nell'accesso a servizi sanitari, e garantendo che i benefici derivanti dai miglioramenti nelle condizioni di vita siano distribuiti in modo più uniforme, specialmente tra gruppi più vulnerabili.

4.4.2 Proiezione della Life Expectancy (e_x).

In questa sezione vengono presentate le proiezioni della Life Expectancy (e_x) a 60 anni, effettuate fino al 2040. L'obiettivo principale è quello di analizzare come le previsioni di questa grandezza di sopravvivenza cambino in funzione del modello statistico utilizzato per la stima. In particolare, si osservano le differenze tra quattro modelli ampiamente utilizzati in letteratura e spiegati nel capitolo 2: Autoregressive Integrated Moving Average (ARIMA), Lee-Carter Model (LC), Cairns-Blake-Dowd Model (CBD, versione M6) e Age-Period-Cohort Model (APC).

Il modello ARIMA si basa sull'analisi delle serie temporali, utilizzando le dinamiche storiche per proiettare il futuro andamento della mortalità. Il modello Lee-Carter, invece, incorpora due dimensioni fondamentali, l'età e il periodo, per analizzare come la mortalità cambi nel tempo a seconda dell'età degli individui. Il modello Cairns-Blake-Dowd (CBD M6) introduce l'effetto coorte, tenendo conto delle variazioni specifiche tra gruppi nati in anni differenti. Infine, il modello Age-Period-Cohort (APC) esamina congiuntamente l'interazione tra età, periodo e coorte, offrendo una visione più articolata dei cambiamenti nelle aspettative di vita.

L'analisi si basa su dati osservati dal 1981 al 2016, mentre le proiezioni vengono realizzate per il periodo dal 2017 al 2040. Il campione di riferimento è costituito da percettori di rendite INPS, con età compresa tra 60 e 100 anni. Le proiezioni sono state effettuate considerando tre diverse combinazioni di caratteristiche degli individui:

- Sesso – categoria pensionistica;
- Sesso – gestione previdenziale;
- Sesso – classe d'importo della pensione;

Per ciascuna combinazione, verranno esaminati i risultati ottenuti con i quattro modelli sopra menzionati, evidenziando le differenze nelle proiezioni dell'aspettativa di vita.

Tale analisi ha diverse applicazioni pratiche. Comprendere come le proiezioni dell'aspettativa di vita variano in funzione del modello utilizzato è fondamentale per diverse ragioni: permette di valutare il rischio di longevità in maniera più accurata, elemento cruciale per le istituzioni previdenziali come l'INPS, che devono assicurare la sostenibilità delle rendite pensionistiche nel lungo periodo. Inoltre, consente di migliorare la pianificazione delle risorse finanziarie, orientando le scelte di policy e gestione dei fondi pensionistici, soprattutto in vista dell'invecchiamento progressivo della popolazione. Infine, offre indicazioni utili per ottimizzare le strategie di investimento previdenziale e per gestire i potenziali impatti economici legati alla crescente Life Expectancy.

4.4.2.1 Combinazione Sesso – Categoria pensionistica.

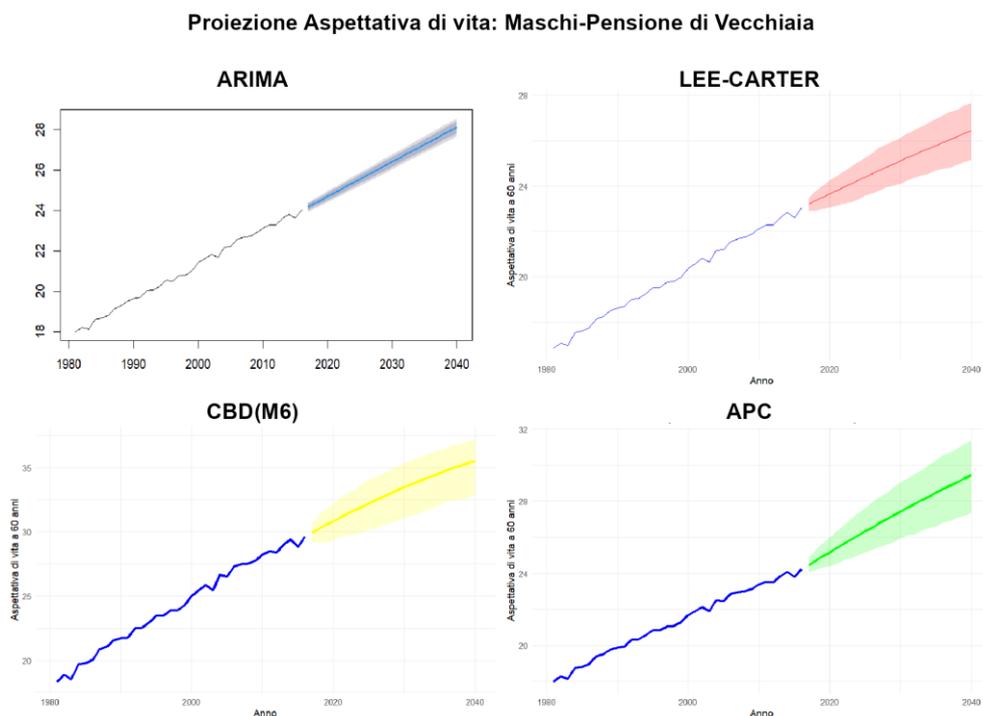


Figura 39: Proiezione Aspettativa di vita: maschi- pensione di vecchiaia.

I grafici presentano le proiezioni dell’aspettativa di vita a 60 anni per i maschi percettori di pensione di vecchiaia, utilizzando quattro modelli statistici: ARIMA, LC, CBD(M6) e APC. Ognuno di questi modelli riflette differenti approcci metodologici e fornisce proiezioni che variano sia in termini di valore assoluto che di incertezza nelle previsioni.

In particolare, nel modello ARIMA, le proiezioni mostrano un costante aumento dell’aspettativa di vita, partendo da circa 24 anni nel 2017 per arrivare a poco più di 27 anni nel 2040. Il trend è lineare e piuttosto stabile, coerente con la natura del modello, che si basa unicamente sui pattern temporali storici. Infatti, come già detto anche in precedenza, il modello ARIMA non considera esplicitamente i fattori legati all’età o al periodo, basandosi unicamente sull’evoluzione storica delle serie temporali. Questo porta a proiezioni meno dinamiche rispetto ad altri modelli che includono variabili strutturali come età, coorte e periodo. Inoltre, l’intervallo di confidenza risulta abbastanza stretto rispetto agli altri modelli, riflettendo una relativa sicurezza nelle previsioni. Ciò implica che il modello stima una crescita dell’aspettativa di vita di 60 anni con una variabilità contenuta, ma potrebbe non catturare eventuali cambiamenti strutturali che influenzano la mortalità in futuro. Questa stabilità del modello è utile quando si desidera fare previsioni con un grado di incertezza ridotto, ma potrebbe mancare di flessibilità rispetto ai modelli che tengono conto delle componenti età- periodo- coorte.

Il modello Lee-Carter è uno dei più utilizzati nel campo della demografia e mostra anch'esso una crescita dell'aspettativa di vita, ma meno pronunciata rispetto al modello ARIMA. Le previsioni indicano che l'aspettativa di vita a 60 anni per i maschi percettori di pensione di vecchiaia raggiunge circa 25-26 anni nel 2040, con un trend di crescita che si stabilizza nel lungo termine. Il modello Lee-Carter è noto per la sua capacità di catturare i cambiamenti storici nel livello di mortalità attraverso un approccio bilanciato tra età e periodo. Tuttavia, la mancanza di un effetto coorte può rendere questo modello meno reattivo a determinati cambiamenti strutturali nel futuro. Quindi, tale modello rappresenta un miglioramento rispetto ad ARIMA perché include l'effetto dell'età, ma è meno sofisticato rispetto a modelli come CBD(M6) o APC. Inoltre, l'ampiezza dell'intervallo di confidenza è maggiore rispetto a quella dell'ARIMA, soprattutto verso la fine del periodo di proiezione. Questo riflette una maggiore incertezza riguardo le previsioni future, il che è plausibile dato che il modello dipende dal trend temporale della mortalità, che può variare più ampiamente nel lungo periodo.

Il modello Cairns-Blake-Dowd (M6) differisce notevolmente dagli altri modelli nelle proiezioni fornite. La Life Expectancy (ex) proiettata per i maschi a 60 anni nel 2040 raggiunge un valore molto elevato, circa 35 anni. Questo implica un significativo incremento rispetto ai modelli ARIMA e Lee-Carter. Infatti, il modello CBD(M6) è particolarmente sensibile all'effetto coorte, che tiene conto delle differenze di mortalità tra le diverse generazioni. In questo caso, l'elevato valore di aspettativa di vita suggerisce che le coorti attualmente invecchianti potrebbero beneficiare di condizioni di sopravvivenza migliori rispetto alle generazioni precedenti. Questo modello è spesso utilizzato per proiezioni a lungo termine proprio perché incorpora l'effetto di coorte, ed è quindi più indicato per analizzare i cambiamenti strutturali che potrebbero influire sulla mortalità. L'interpretazione di queste proiezioni più alte potrebbe derivare da progressi in ambito sanitario e socioeconomico che influenzano specificamente questa coorte di maschi anziani. In questo caso però, si osserva un intervallo di confidenza più ampio, riflettendo una maggiore incertezza associata a queste proiezioni ottimistiche.

Infine, il modello APC mostra una crescita dell'aspettativa di vita simile al Lee-Carter, con una proiezione che raggiunge circa 25-26 anni nel 2040. Tuttavia, il modello APC ha il vantaggio di considerare congiuntamente l'età, il periodo e la coorte, offrendo una rappresentazione più completa dei fattori che influenzano la mortalità. Come detto, il modello APC cattura simultaneamente l'effetto dell'età, del periodo (cambiamento a livello storico) e l'effetto coorte (diverse generazioni che sperimentano condizioni di sopravvivenza differenti). Grazie a questa struttura, il modello APC è in grado di adattarsi meglio a cambiamenti demografici più complessi, sebbene la sua complessità possa portare a maggiore incertezza nelle previsioni a lungo termine. Infatti, anche in questo caso, l'intervallo di confidenza diventa più ampio man mano che ci si avvicina al 2040, indicando una crescente incertezza. Questo è in parte dovuto alla complessità del modello,

che tenta di bilanciare più componenti, ognuna delle quali può contribuire a una maggiore variabilità nelle proiezioni.

In conclusione, queste differenze tra i modelli suggeriscono che, mentre i metodi più semplici (come ARIMA) possono fornire proiezioni più stabili, modelli più complessi (come CBD e APC) tendono a incorporare maggiori dettagli sulle dinamiche della mortalità, ma al prezzo di una maggiore incertezza nelle stime future. Ciò è particolarmente rilevante per le istituzioni previdenziali come l'INPS, che devono considerare non solo l'aspettativa di vita media, ma anche la variabilità e l'incertezza associata alle previsioni di mortalità.

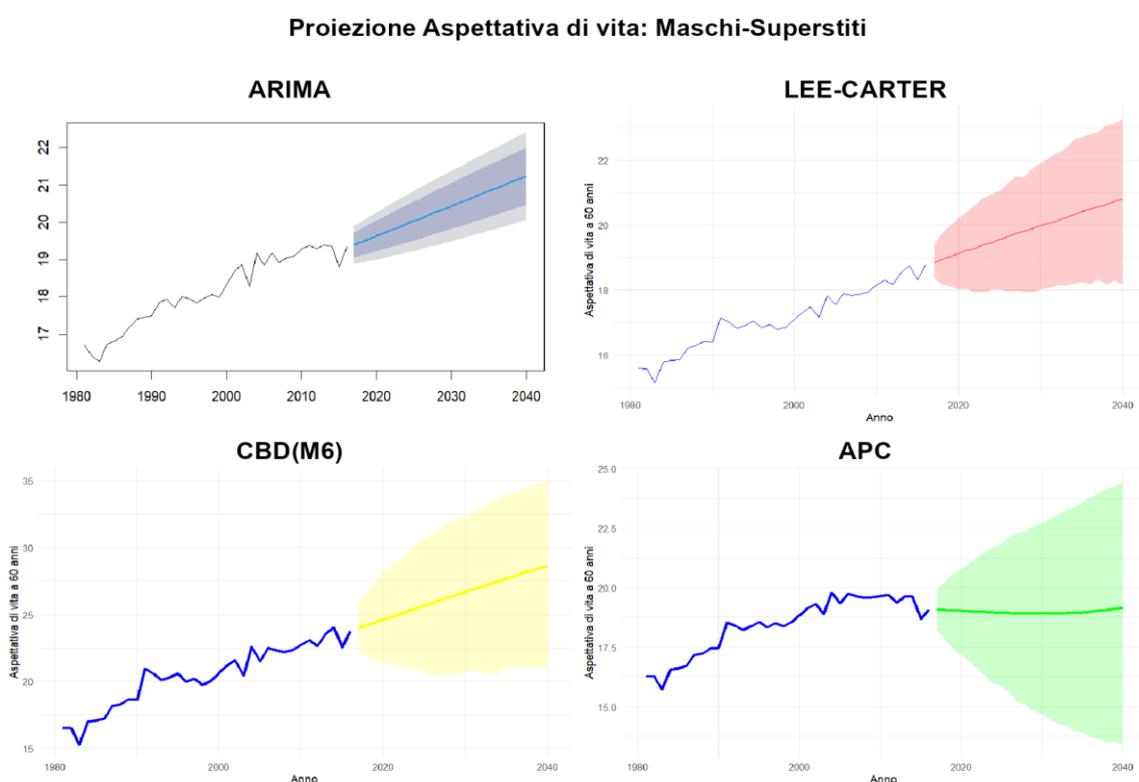


Figura 40: Proiezione aspettativa di vita: Maschi, percettori di pensione ai superstiti.

La Figura 40, fa riferimento alla proiezione dell'aspettativa di vita dei maschi percettori di pensione ai superstiti.

Il modello ARIMA prevede un incremento dell'aspettativa di vita dai 19 anni nel 2017 fino a circa 21,5 anni nel 2040. Rispetto alla combinazione maschi-vecchi, l'andamento qui è meno pronunciato, suggerendo una crescita più moderata. Inoltre, si osserva una maggiore incertezza rispetto al caso precedente, come dimostrato dall'intervallo di confidenza più ampio. Ciò potrebbe indicare che il modello ARIMA ha difficoltà a prevedere in modo stabile l'aspettativa di vita per questo gruppo specifico, dove le caratteristiche demografiche e di mortalità possono variare maggiormente.

Il modello Lee-Carter prevede un'aspettativa di vita a 60 anni per i maschi superstiti che si stabilizza intorno ai 22 anni nel 2040. Anche in questo caso, la crescita è meno marcata rispetto al gruppo dei maschi-vecchi, con un andamento più graduale. L'ampiezza dell'intervallo di confidenza è notevolmente più ampia rispetto al caso della pensione di vecchiaia, soprattutto a partire dal 2025, suggerendo che le proiezioni per i superstiti sono soggette a una maggiore variabilità. Ciò potrebbe essere legato alla natura più eterogenea di questa categoria, che include individui con diverse caratteristiche di sopravvivenza e condizioni sociali ed economiche.

Anche in questo caso, il modello CBD(M6) mostra un'aspettativa di vita superiore a quella stimata dagli altri modelli, con una proiezione che arriva a circa 25 anni nel 2040. Tuttavia, rispetto alla combinazione maschi-vecchi, la crescita è meno pronunciata, il che potrebbe riflettere la diversa struttura di mortalità per i percettori di pensione ai superstiti, ovvero i vedovi. Anche qui, l'intervallo di confidenza è ampio, indicando una maggiore incertezza. Questo potrebbe essere dovuto all'inclusione dell'effetto coorte che, sebbene utile per catturare le dinamiche generazionali, introduce una certa volatilità nelle previsioni a lungo termine per questo gruppo di pensionati.

Infine, il modello APC proietta un'aspettativa di vita più stabile rispetto agli altri modelli, con un valore che si assesta poco sopra i 20 anni nel 2040. Questa stabilizzazione potrebbe essere interpretata come il risultato di due fattori principali. Da un lato, il fattore età e il fattore periodo sembrano avere un'influenza più contenuta sulla sopravvivenza dei maschi superstiti, dall'altro lato, il fattore coorte potrebbe indicare che le generazioni attuali e future di superstiti non beneficiano più dei miglioramenti sanitari e socioeconomici che invece hanno caratterizzato le generazioni precedenti. L'intervallo di confidenza associato alle previsioni del modello APC è piuttosto ampio e aumenta notevolmente con l'avanzare del tempo, specialmente a partire dal 2025. Questo suggerisce una crescente incertezza riguardo le proiezioni nel lungo periodo. La stabilizzazione della curva a circa 20,5 anni combinata con un'ampia incertezza indica che, sebbene il modello preveda una relativa stagnazione dell'aspettativa di vita, vi sono molte variabili esterne che potrebbero alterare significativamente le condizioni future per i maschi superstiti.

Proiezione Aspettativa di vita: Femmine-Pensione di Vecchiaia

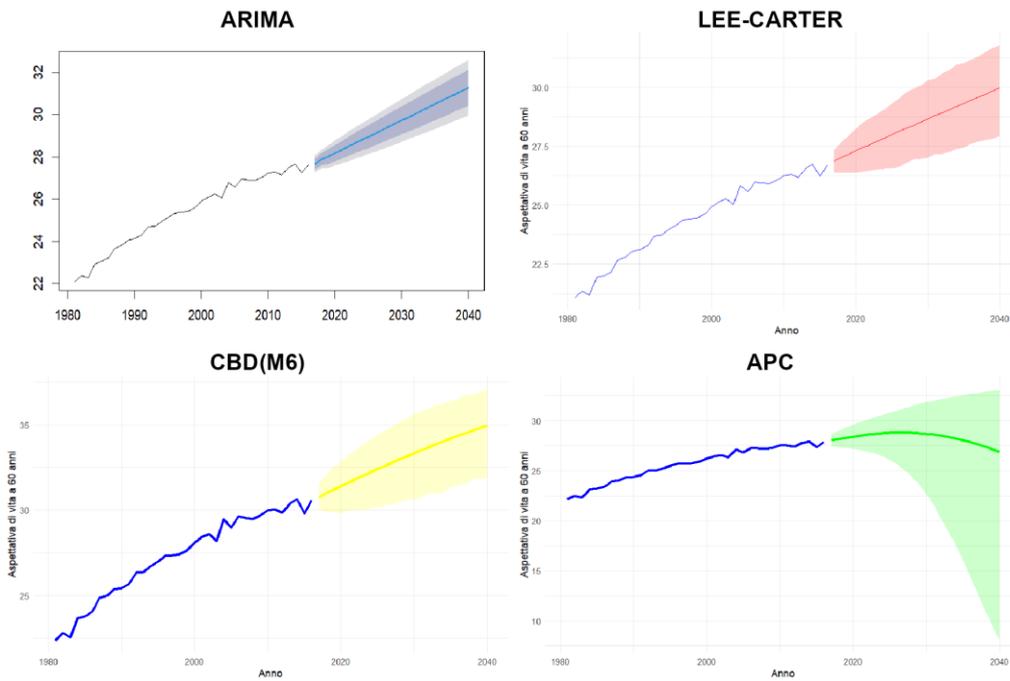


Figura 41:Proiezione aspettativa di vita: femmine-pensione di vecchiaia.

Proiezione Aspettativa di vita: Femmine-Superstiti

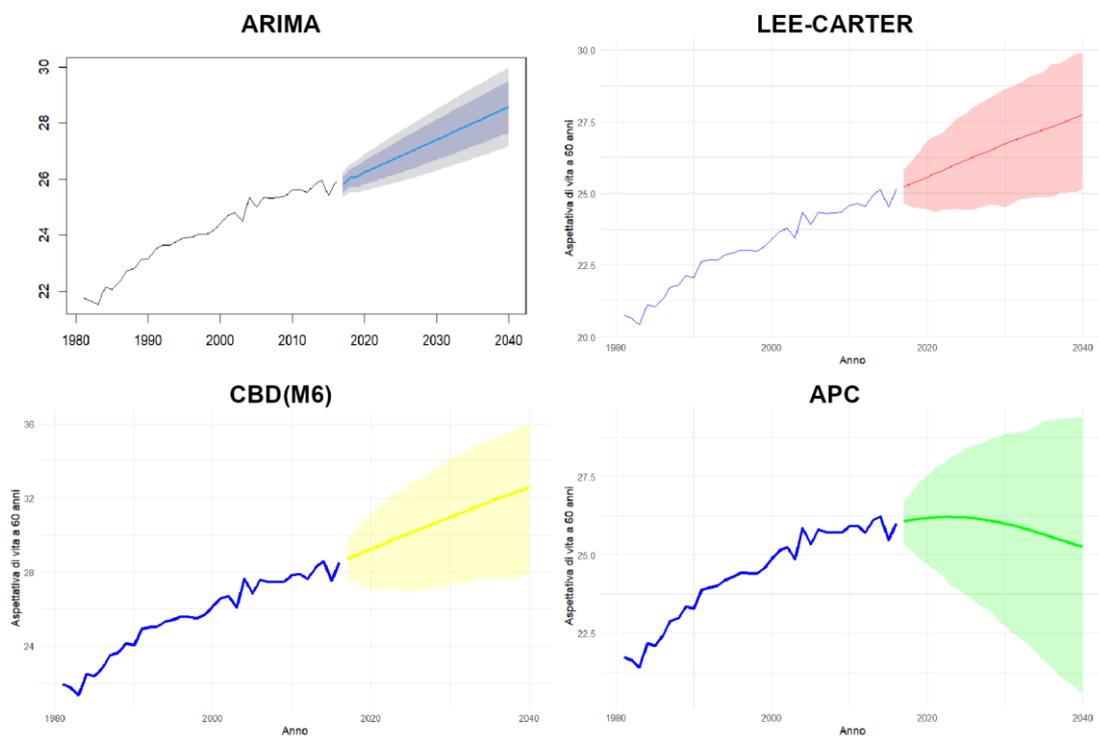


Figura 42:Proiezione aspettativa di vita: femmine- percettrici di pensione ai superstiti.

La Figura 41 e la Figura 42 mostrano rispettivamente le proiezioni per l'aspettativa di vita per le donne percettrici di pensione di vecchiaia e pensione ai superstiti. Tali proiezioni mostrano andamenti simili nei vari modelli, ma con alcune differenze in termini di livello assoluto e di incertezza delle previsioni. In entrambi i casi, le proiezioni confermano quanto visto nei paragrafi precedenti, ovvero che le donne tendono ad avere un'aspettativa di vita più alta rispetto agli uomini, riflettendo il noto vantaggio della longevità femminile. Tuttavia, ci sono alcune differenze significative tra le due categorie e tra i diversi modelli utilizzati.

In entrambi i gruppi, il modello ARIMA prevede un incremento moderato dell'aspettativa di vita, con valori che salgono dai 26-27 anni nel 2017 circa 30-31 nel 2040. Per le femmine percettrici di pensione di vecchiaia il modello prevede una crescita più lineare e stabile, raggiungendo i 31 anni nel 2040. Per le femmine superstiti, l'aspettativa di vita è leggermente più bassa e cresce fino ai 29 anni nel 2040. L'intervallo di confidenza si allarga verso il 2040, ma rimane abbastanza contenuto rispetto agli altri modelli, indicando una certa affidabilità delle proiezioni nel lungo termine. Bisogna però ricordare che il modello ARIMA non tiene conto di fattori strutturali come l'effetto coorte, il che può limitare la sua capacità di adattarsi a cambiamenti futuri più complessi.

Il modello Lee-Carter mostra una crescita più marcata dell'aspettativa di vita per entrambe le categorie, anche se rimane sotto i livelli proiettati dal modello CBD. Le femmine con pensione di vecchiaia arrivano a circa 30 anni di aspettativa di vita nel 2040, invece nello stesso anno le femmine superstiti arrivano ad un valore vicino ai 29 anni. L'intervallo di confidenza nel modello Lee-Carter è ampio, suggerendo una maggiore incertezza. Questo è legato alla sensibilità del modello agli andamenti temporali della mortalità, che potrebbero variare in modo imprevisto.

Il modello CBD(M6), anche in questo caso, è quello che fornisce le proiezioni più ottimistiche per entrambe le categorie. Per le femmine percettrici di pensione di vecchiaia, l'aspettativa di vita raggiunge quasi 35 anni nel 2040, partendo da circa 29 anni nel 2017. Mentre le femmine superstiti arrivano ad un'aspettativa di 33 anni nel 2040. L'ottimismo legato a questo modello, suggerisce che le donne continueranno a beneficiare di miglioramenti delle condizioni di vita e sanitarie nel futuro.

A questo punto, si osserva un andamento particolare, legato al modello APC. Infatti, tale modello mostra una tendenza alla stabilizzazione dell'aspettativa di vita per entrambe le categorie a partire dalla metà degli anni '20. Per le femmine percettrici di pensione di vecchiaia la stabilizzazione avviene intorno ai 30 anni, mentre per le superstiti intorno ai 27,5 anni. Questo invece, potrebbe suggerire che i miglioramenti in termini di longevità femminile potrebbero non proseguire allo stesso ritmo osservato in passato.

Dall'analisi delle proiezioni dell'aspettativa di vita per la combinazione sesso-categoria pensionistica, emerge una chiara differenza tra i percettori di pensioni di vecchiaia e i

percettori di pensioni ai superstiti, con le donne che tendono ad avere un'aspettativa di vita più elevata rispetto agli uomini in entrambe le categorie. I modelli più complessi, come il CBD (M6), suggeriscono una crescita più marcata dell'aspettativa di vita, mentre il modello APC prevede una stabilizzazione nel lungo periodo, particolarmente per le categorie di superstiti. Queste proiezioni evidenziano l'importanza della scelta del modello nella valutazione della longevità, con implicazioni dirette per la sostenibilità dei sistemi pensionistici, dove l'incertezza nelle previsioni a lungo termine gioca un ruolo significativo.

4.4.2.2 Combinazione sesso-gestione previdenziale.

A questo punto, si va ad effettuare una valutazione della proiezione dell'aspettativa di vita, effettuata con i 4 modelli sopracitati, della combinazione sesso – gestione previdenziale:

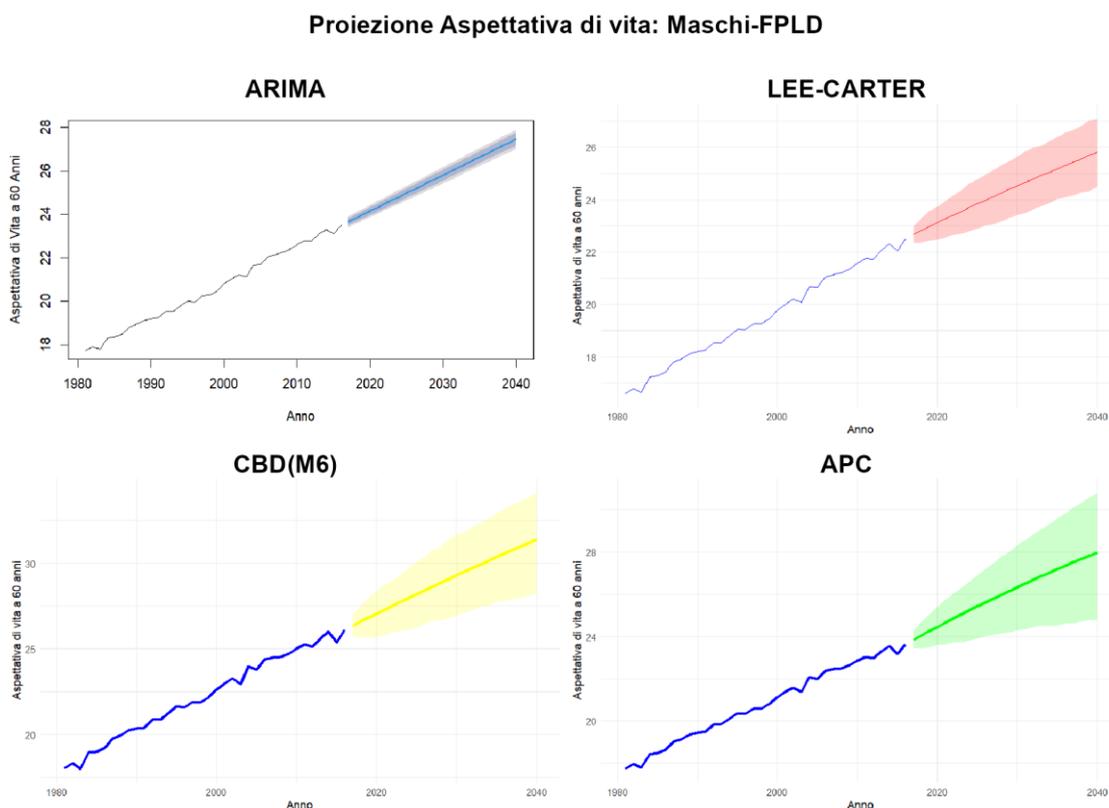


Figura 43: Proiezione aspettativa di vita: maschi appartenenti al Fondo Pensione Lavoratori Dipendenti

Proiezione Aspettativa di vita: Maschi-Coltivatori diretti, Coloni e Mezzadri

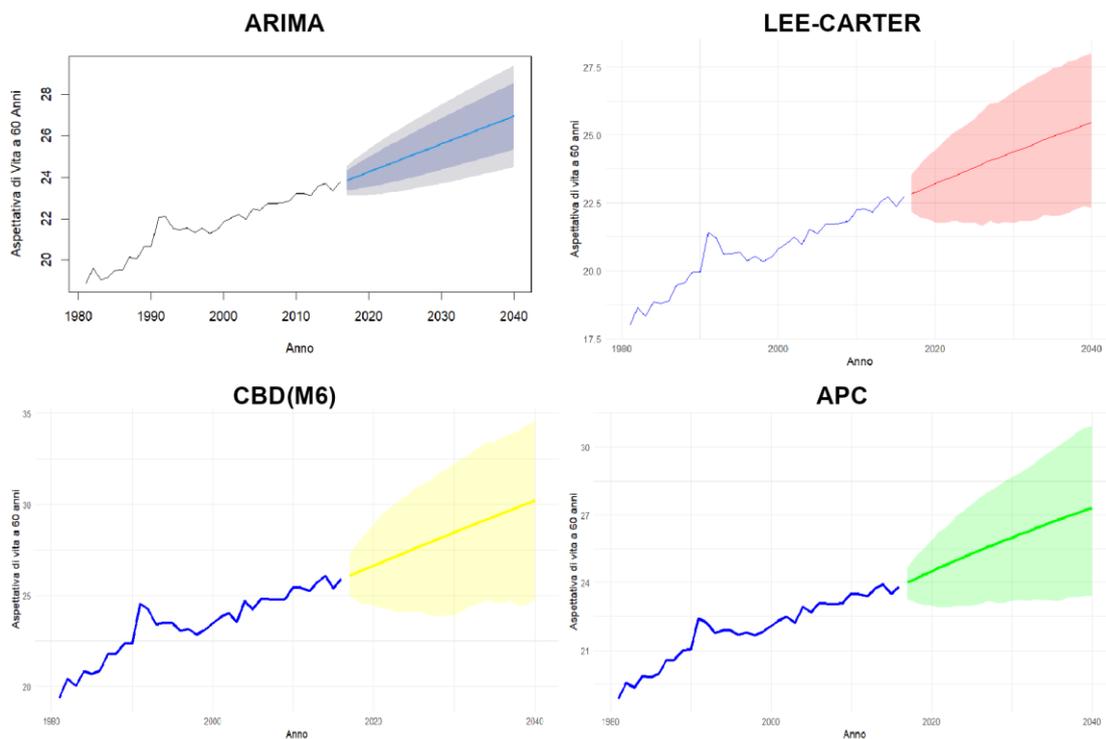


Figura 44: Proiezione aspettativa di vita: maschi, coltivatori diretti, coloni e mezzadri.

Proiezione Aspettativa di vita: Maschi-Artigiani e Commercianti

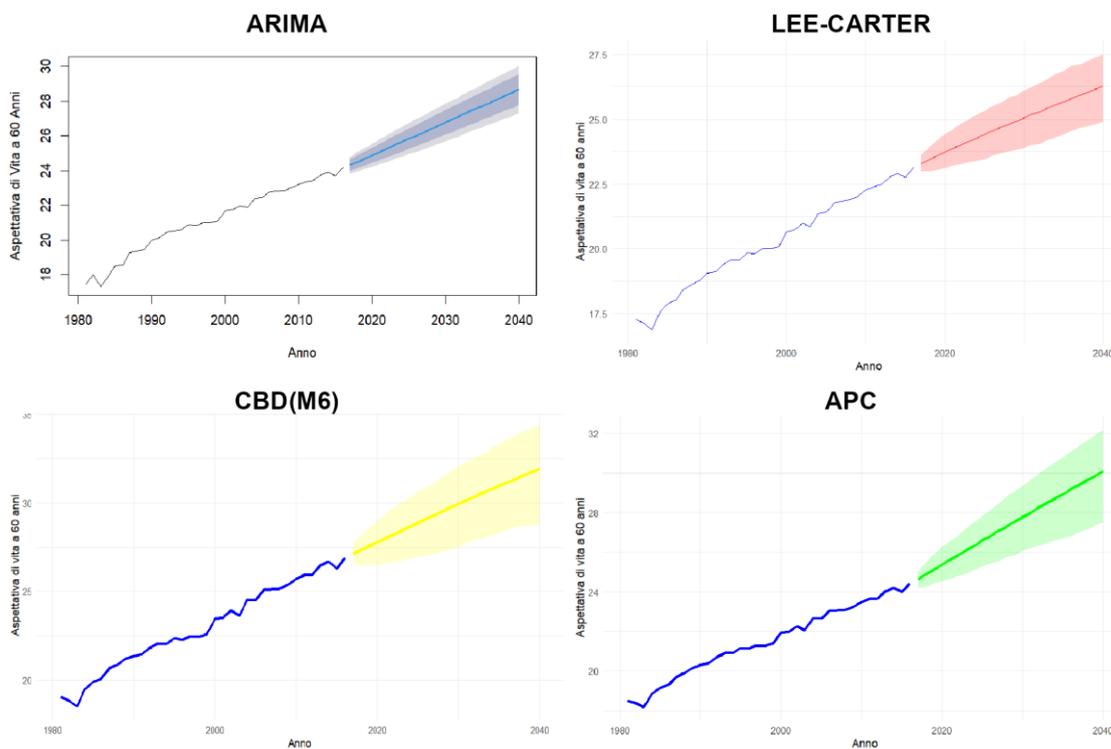


Figura 45: Proiezione aspettativa di vita: Maschi, Artigiani e Commercianti.

Le Figure 43, 44 e 45 rappresentano i grafici dell'aspettativa di vita per i maschi appartenenti al FPLD, ai Coltivatori diretti, coloni e Mezzadri, e agli Artigiani e Commercianti. Tali grafici rilevano alcune importanti differenze nelle dinamiche di sopravvivenza tra queste categorie, a seconda del modello utilizzato. In generale, si osserva una tendenza alla crescita dell'aspettativa di vita in tutte le categorie, anche se i livelli assoluti e le traiettorie di crescita variano a seconda della gestione previdenziale e del modello scelto.

Per quanto riguarda i maschi appartenenti al FPLD, i modelli mostrano una crescita costante e progressiva dell'aspettativa di vita, con alcune differenze nelle previsioni finali. Il modello ARIMA suggerisce una proiezione lineare con un'aspettativa di vita che supera leggermente i 27 anni entro il 2040. Il modello Lee-Carter, invece, prevede una crescita leggermente più contenuta, con un valore finale attorno ai 26 anni. D'altra parte, il modello CBD(M6) fornisce ancora una volta previsioni più ottimistiche, stimando un'aspettativa di vita che si avvicina ai 30 anni entro il 2040. Questo riflette una maggiore sensibilità del modello all'effetto coorte, che prevede un miglioramento delle condizioni di sopravvivenza per le generazioni future. Il modello APC, pur presentando una traiettoria crescente, mostra una previsione più intermedia rispetto al CBD, stabilizzandosi intorno ai 28 anni. È interessante notare che i margini di incertezza, rappresentati dagli intervalli di confidenza, sono relativamente contenuti nel caso ARIMA, ma si ampliano considerevolmente nei modelli Lee-Carter, CBD e APC, segnalando che le previsioni a lungo termine per i maschi appartenenti al FPLD sono soggette a una crescente incertezza, in particolare in modelli più complessi.

Nel caso dei Coltivatori diretti, Coloni e Mezzadri, le proiezioni dell'aspettativa di vita mostrano un andamento simile, ma con livelli assoluti leggermente inferiori rispetto ai maschi appartenenti al FPLD. Questo potrebbe riflettere le condizioni socioeconomiche meno favorevoli di questo gruppo, che generalmente gode di un'aspettativa di vita inferiore. Il modello ARIMA prevede una crescita fino a circa 26 anni entro il 2040, mentre il modello Lee-Carter proietta un'aspettativa di vita che raggiunge i 27,5 anni nello stesso periodo. Ancora una volta il modello CBD(M6) fornisce previsioni molto più ottimistiche, con l'aspettativa di vita che arriva a superare i 30 anni entro il 2040 mostrando una maggiore attenzione all'effetto coorte. Infine, il modello APC prevede una crescita più modesta, con un'aspettativa di vita che raggiunge poco più di 28 anni, anche se, come nel caso del FPLD, l'incertezza nelle previsioni a lungo termine è piuttosto elevata.

Infine, per quanto riguarda gli Artigiani e Commercianti, i modelli presentano una traiettoria simile a quella dei Coltivatori, con livelli di aspettativa di vita che, tuttavia, risultano leggermente più elevati. Il modello ARIMA proietta un'aspettativa di vita di circa 29 anni entro il 2040, mentre il Lee-Carter prevede un valore leggermente inferiore, intorno ai 27,5 anni. Il modello CBD (M6) si distingue nuovamente per le sue previsioni più ottimistiche, stimando un'aspettativa di vita di circa 32 anni nel 2040. Questo modello

riflette una visione ottimistica, soprattutto grazie all'inclusione dell'effetto coorte, che suggerisce che le nuove generazioni di artigiani e commercianti potrebbero beneficiare di condizioni di sopravvivenza migliori rispetto alle precedenti. Il modello APC, invece, segue una traiettoria più moderata, stabilizzandosi attorno ai 30 anni, con un'ampia incertezza a partire dal 2025.

I modelli ARIMA e Lee-Carter forniscono previsioni più conservative e tendono a presentare incrementi costanti e gradualmente dell'aspettativa di vita. Al contrario, il modello CBD (M6) è il più ottimistico in tutte le categorie, grazie all'inclusione dell'effetto coorte, che lo rende particolarmente sensibile ai miglioramenti generazionali delle condizioni di vita. Il modello APC, pur prevedendo una crescita simile, suggerisce una stabilizzazione a lungo termine per tutte le categorie, accompagnata da una notevole incertezza nelle proiezioni future. Tale incertezza nelle proiezioni a lungo termine rimane un tema centrale, soprattutto per i modelli Lee-Carter, CBD e APC, che evidenziano un'ampia variabilità nelle previsioni future. Questo potrebbe riflettere l'incertezza intrinseca nelle dinamiche di mortalità per gruppi diversi e la difficoltà nel prevedere come fattori esterni, quali le politiche sanitarie e i miglioramenti nelle condizioni socioeconomiche, influenzeranno le varie categorie di lavoratori nel lungo periodo.

Proiezione Aspettativa di vita: Femmine-FPLD

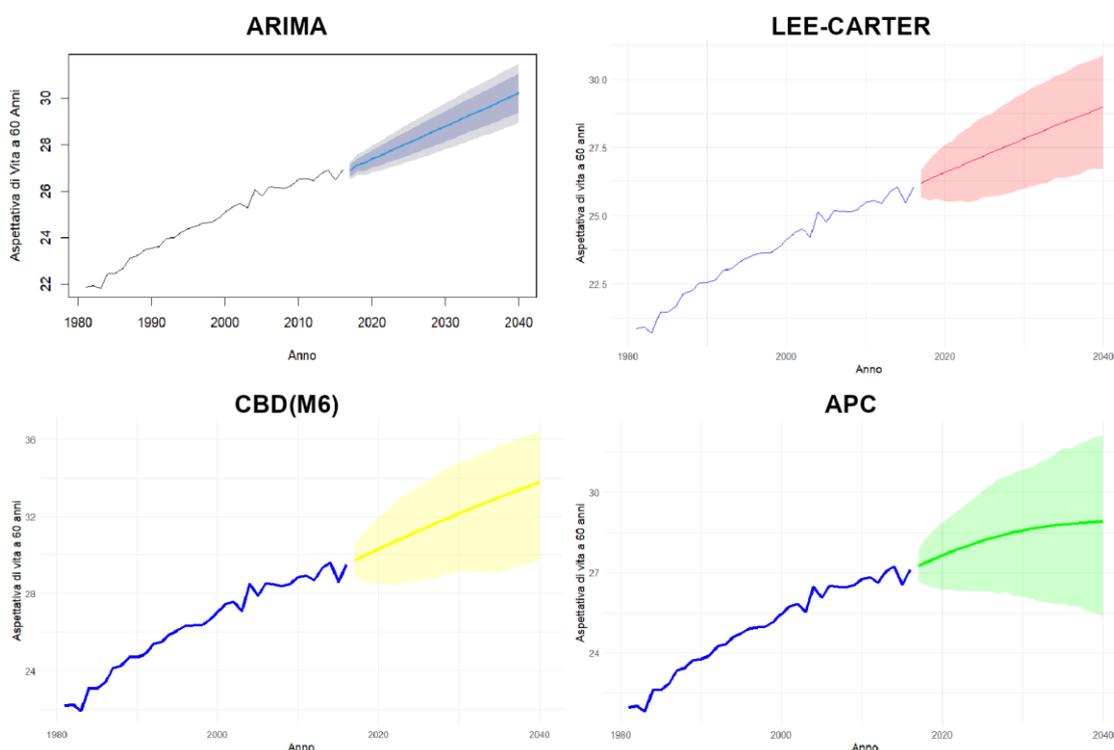


Figura 46: Proiezione aspettativa di vita: Femmine appartenenti al Fondo Pensione Lavoratori Dipendenti.

Proiezione Aspettativa di vita: Femmine- Coltivatori diretti, Coloni e Mezzadri

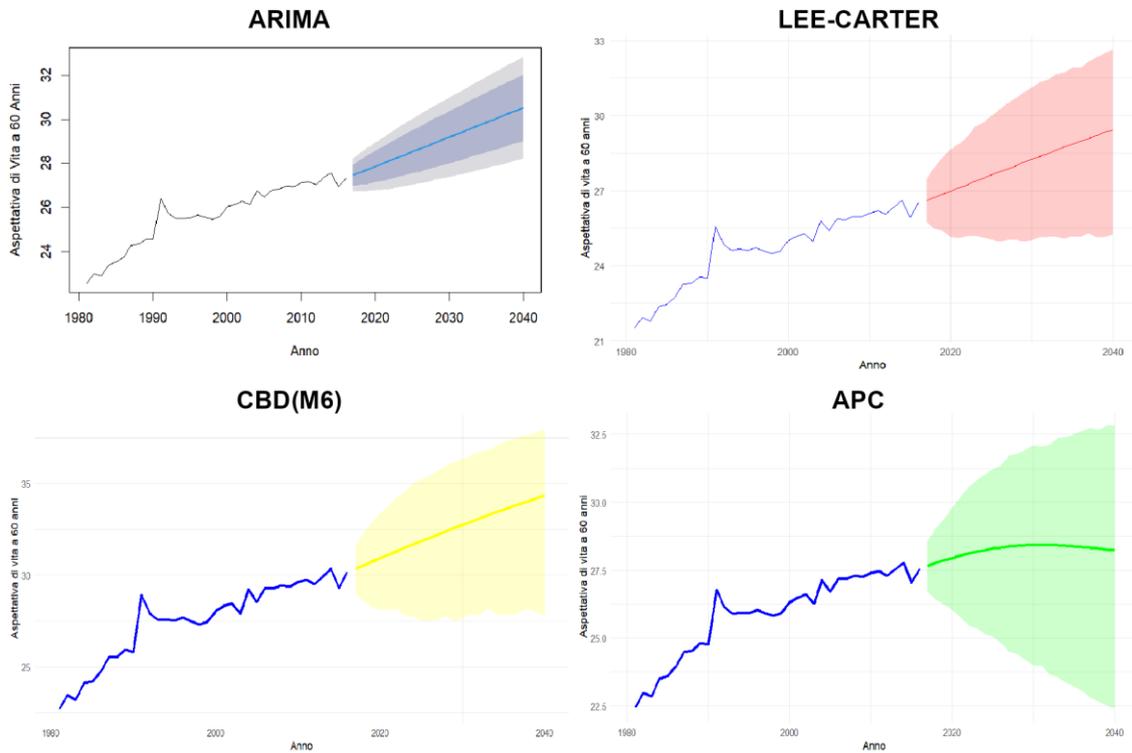


Figura 47: Proiezione aspettativa di vita: Femmine, Coltivatori diretti, coloni e mezzadri.

Proiezione Aspettativa di vita: Femmine-Artigiani e Commercianti

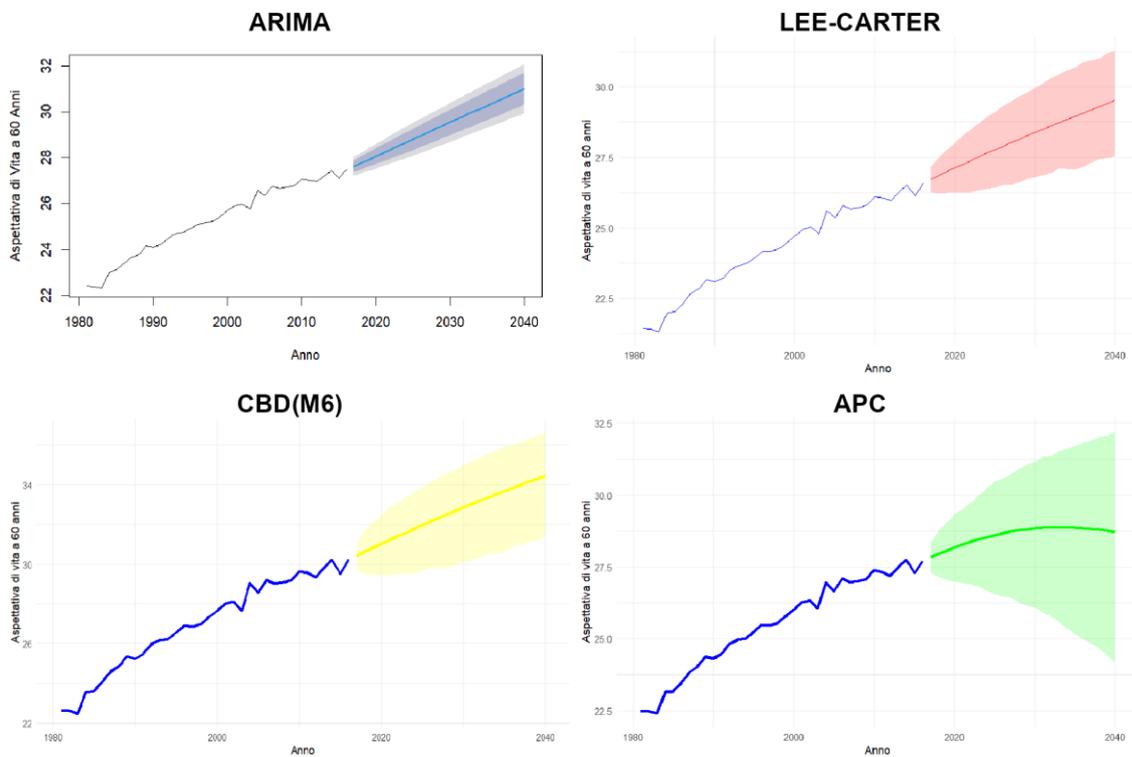


Figura 48: Proiezione aspettativa di vita: Femmine appartenenti alla gestione Artigiani e commercianti.

Le proiezioni dell'aspettativa di vita per le donne (osservate nelle Figure 46, 47 e 48) appartenenti alle tre diverse gestioni previdenziali (FPLD, Coltivatori diretti, Coloni e Mezzadri, Artigiani e Commercianti) mostrano un andamento generalmente crescente in tutte le categorie, sebbene vi siano importanti differenze sia nei livelli assoluti che nella traiettoria prevista dai diversi modelli. È evidente che, come già emerso in altre analisi, le donne presentano in media un'aspettativa di vita più alta rispetto agli uomini, ma la crescita prevista varia significativamente in base al modello considerato e alla specifica gestione previdenziale.

Il modello ARIMA prevede un'aspettativa di vita che cresce in modo lineare e costante in tutte e tre le categorie. Per le donne appartenenti al FPLD, l'aspettativa di vita parte da circa 27 anni nel 2017 e arriva a superare i 30 anni nel 2040. Anche per le donne Artigiane e Commercianti, l'andamento è simile, con un valore di partenza leggermente superiore (circa 28 anni) e una crescita che porta a 31 anni nel 2040. Per le donne appartenenti alla gestione dei Coltivatori diretti, Coloni e Mezzadri, l'aspettativa di vita è invece leggermente più bassa, partendo da 27,5 anni nel 2020 e raggiungendo i 30 anni nel 2040. Sebbene il modello offra proiezioni più lineari e conservative, gli intervalli di confidenza si ampliano verso il 2040, riflettendo una maggiore incertezza nelle previsioni a lungo termine, anche per una crescita che appare costante.

Il modello Lee-Carter, che tiene conto del periodo e dell'età, prevede una crescita leggermente più contenuta rispetto all'ARIMA, ma con intervalli di confidenza significativamente più ampi. Per le donne appartenenti al FPLD, l'aspettativa di vita cresce fino a circa 30 anni nel 2040. La tendenza è simile per le donne artigiane e commercianti, che raggiungono anch'esse circa i 30 anni alla fine del periodo di proiezione, mentre le donne coltivatrici vedono una crescita leggermente superiore che raggiunge circa i 31 anni nel 2040. L'ampio intervallo di confidenza, in particolare dopo il 2025, indica una maggiore incertezza riguardo l'andamento della mortalità futura. Il risultato è una crescita dell'aspettativa di vita che, pur continuando a migliorare, appare più prudente, con una variabilità più ampia man mano che ci si avvicina al 2040.

Il modello CBD (M6), anche in questo caso, è quello che offre le proiezioni più ottimistiche per tutte le categorie. Infatti, per le donne dipendenti l'aspettativa di vita potrebbe raggiungere i 35 anni nel 2040, per le donne artigiane e commercianti si arriva a circa 34 anni, mentre per le donne appartenenti al settore agricolo, i valori ci si aspetta che raggiungano i 33 anni circa alla fine del periodo di proiezione. Però, nonostante queste proiezioni siano molto ottimistiche, l'ampiezza degli intervalli di confidenza è significativa, riflettendo un'elevata incertezza, specialmente verso la fine del periodo di proiezione. Questo indica che, sebbene il modello preveda miglioramenti generazionali marcati, tali proiezioni devono essere interpretate con cautela, in quanto si basano sull'assunzione che le nuove coorti sperimentano miglioramenti sostanziali rispetto alle precedenti.

Infine, il modello APC, che incorpora l'interazione tra età, periodo e coorte, prevede una crescita più contenuta rispetto al CBD. Infatti, in questo caso per tutte e tre le gestioni previdenziali oggetto di studio al quale appartengono le donne, l'aspettativa tende a stabilizzarsi tra i 29 e i 30 anni. Inoltre, anche per il modello APC c'è un ampio intervallo di confidenza che si allarga notevolmente verso la fine del periodo di proiezione. Ciò significa che man mano ci si sposta in avanti con il tempo, più l'incertezza legata alle proiezioni aumenta. Ciò è particolarmente evidente per le categorie più svantaggiate, come i Coltivatori, che potrebbero non beneficiare in egual misura dei miglioramenti sanitari e socioeconomici. Il modello APC offre una visione più prudente rispetto al CBD, suggerendo che, sebbene l'aspettativa di vita continui a crescere, potrebbe stabilizzarsi nel lungo termine, con una maggiore variabilità dovuta ai cambiamenti periodici e generazionali.

In tutti i modelli, l'ampiezza degli intervalli di confidenza aumenta considerevolmente nel lungo termine, segnalando che le previsioni a lungo termine sono soggette a notevole incertezza, soprattutto per le categorie meno avvantaggiate, come i Coltivatori diretti. Questa incertezza legata a coloro che effettuano lavori agricoli può essere dovuta a diversi fattori: lavoro agricolo esposto a rischi fisici e ambientali; progressi non uniformi nelle generazioni; modernizzazione agricola variabile. Questo rende più difficile prevedere come evolveranno le loro condizioni di vita e, di conseguenza, la mortalità a lungo termine. Tutto ciò genera intervalli di confidenza più ampi nelle proiezioni, soprattutto per modelli complessi come il CBD(M6) e l'APC. Questo lo si può osservare sia per gli uomini che per le donne.

4.4.2.3 Combinazione sesso - classe d'importo della pensione.

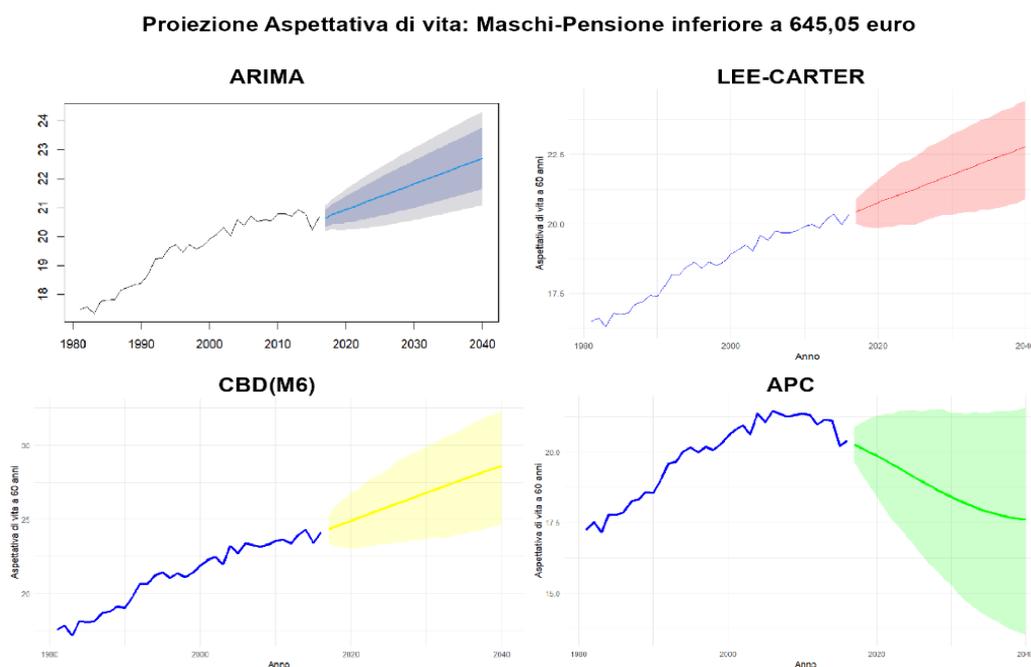


Figura 49:Proiezione aspettativa di vita: Maschi con pensione inferiore a 645,05 euro.

Proiezione Aspettativa di vita: Maschi-Pensione compresa tra 645,05 e 1290,10 euro

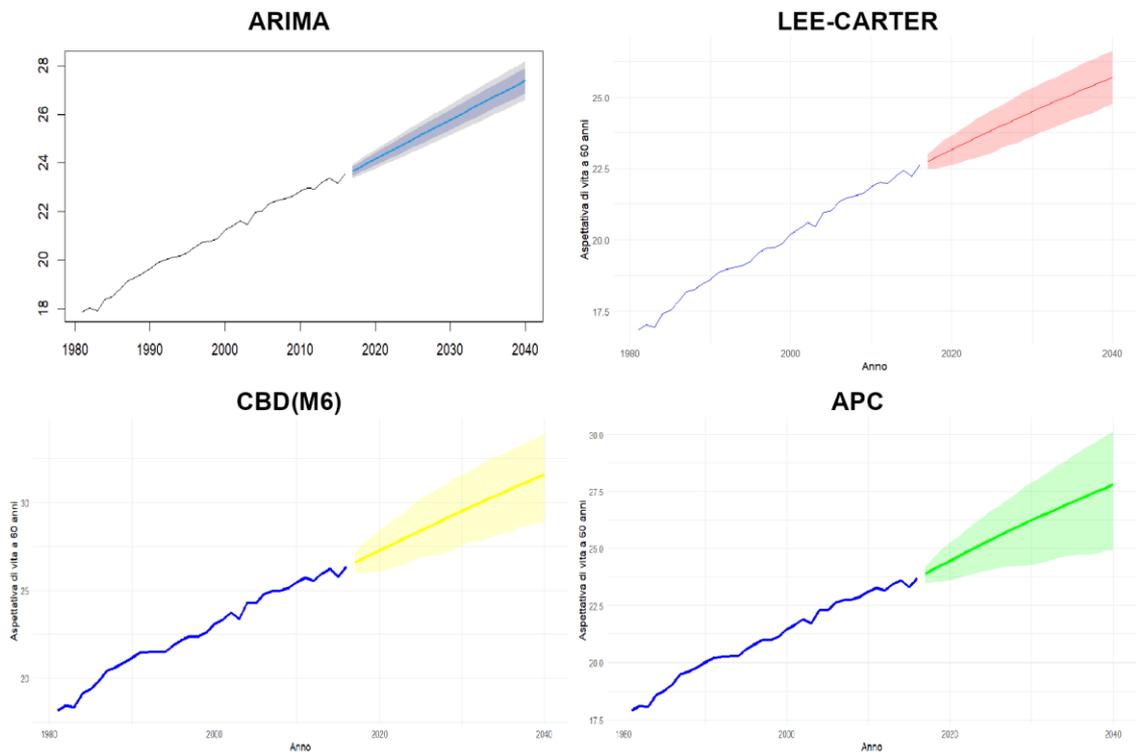


Figura 50: Proiezione aspettativa di vita: Maschi con pensione compresa tra 645,05 e 1290,10 euro.

Proiezione Aspettativa di vita: Femmine-Pensione inferiore a 645,05 euro

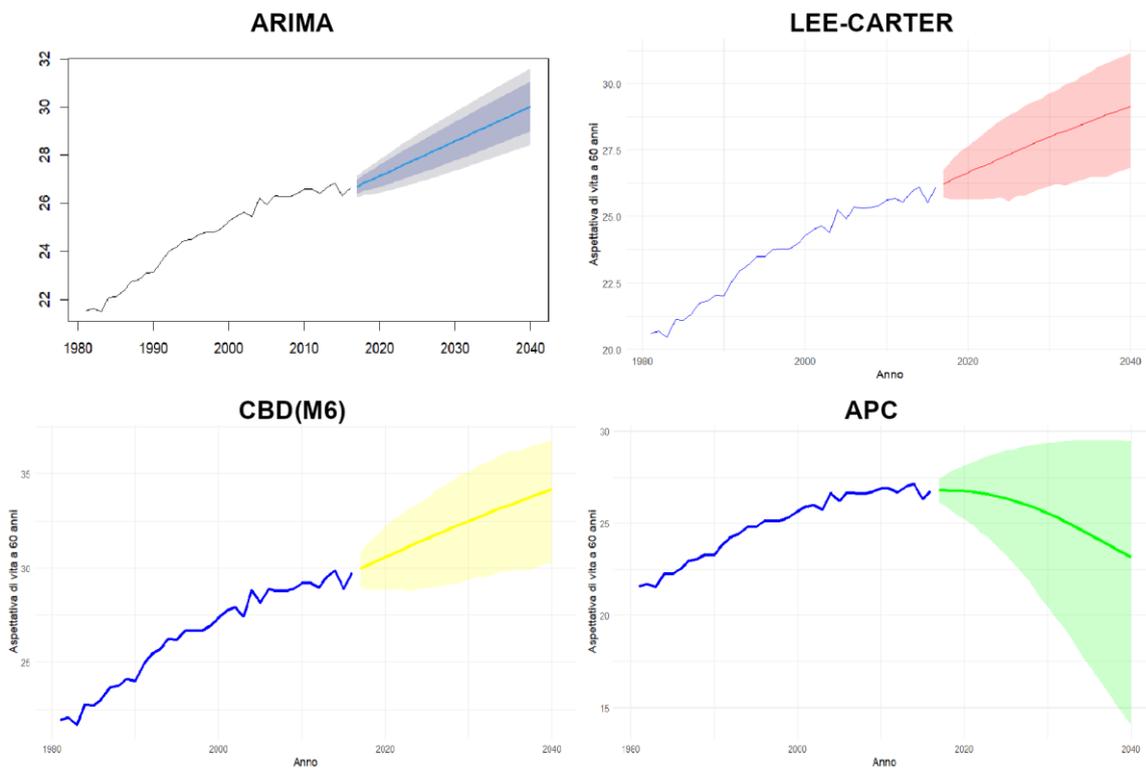


Figura 51: Proiezione aspettativa di vita: Femmine con pensione inferiore a 645,05 euro.

Proiezione Aspettativa di vita: Femmine-Pensione compresa tra 645,05 e 1290,10 euro

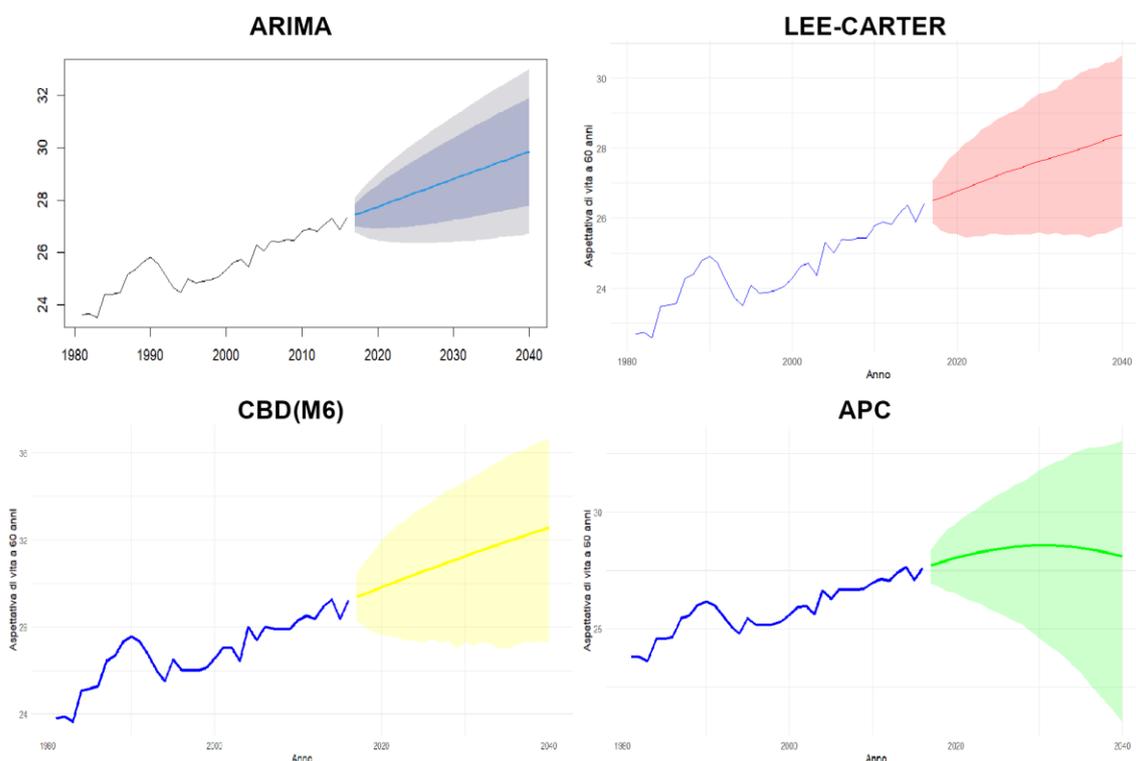


Figura 52: Proiezione aspettativa di vita: Femmine con pensione compresa tra 645,05 e 1290,10 euro.

Le Figure 49, 50, 51 e 52 rappresentano le proiezioni dell'aspettativa di vita per i maschi e per le femmine con pensioni inferiori a 645,05 euro e per quelli con pensioni comprese tra 645,05 e 1.290,10 euro. Tali rappresentazioni rivelano alcune differenze significative, che possono essere analizzate considerando sia le combinazioni pensionistiche sia i diversi modelli statistici utilizzati per le previsioni.

Le differenze osservate nelle proiezioni tra le diverse classi d'importo derivano principalmente dalle condizioni socioeconomiche degli individui che percepiscono le pensioni. In generale:

- Gli individui con pensioni inferiori a 645,05 euro tendono a mostrare aspettative di vita più basse rispetto a quelli con pensioni più elevate. Ciò è coerente con le tendenze ben note nella letteratura demo-previdenziale, dove livelli di reddito più bassi sono spesso associati a condizioni di vita meno favorevoli e, di conseguenza, a una mortalità più elevata. Questi individui potrebbero aver sperimentato condizioni lavorative più svantaggiate, con un accesso limitato a servizi sanitari, stili di vita meno salutari, e una minore possibilità di mantenere buone condizioni di salute durante la vecchiaia.
- Gli individui che percepiscono una pensione compresa tra 645,05 e 1.290,10 euro, ed in particolare le donne, tendono ad avere un'aspettativa di vita più elevata rispetto alla categoria con pensioni inferiori. Questo riflette una maggiore stabilità

economica e, probabilmente, un accesso migliore alle cure sanitarie e a condizioni di vita più favorevoli.

Invece, analizzando le differenze tra i modelli si osserva come il modello ARIMA proietti una crescita lineare e stabile dell'aspettativa di vita per entrambe le categorie pensionistiche. Tuttavia, nel periodo di proiezione, gli individui con pensioni più basse mostrano livelli più ridotti di aspettativa di vita a 60 anni rispetto a quelli con pensioni più alte. Per esempio, per i maschi con pensioni inferiori a 645,05 euro, l'aspettativa di vita proiettata per il 2040 è inferiore ai 24 anni, mentre per i maschi con pensioni tra 645,05 e 1.290,10 euro, l'aspettativa di vita prevista è di circa 27 anni. Lo stesso schema si ripete per le femmine, con aspettative di vita più elevate, ma differenze simili tra le categorie di reddito. L'ARIMA riflette trend storici stabili, ma non incorpora variabili come l'effetto coorte o fattori socioeconomici specifici, il che lo rende un modello con previsioni più conservative e regolari.

Il modello Lee-Carter mostra anch'esso una tendenza di crescita, ma con un intervallo di confidenza più ampio rispetto ad ARIMA. Anche qui è dimostrato che per gli individui con pensioni più basse, l'aspettativa di vita è significativamente inferiore rispetto a quelli con pensioni più alte. Questo modello tiene conto dell'effetto temporale, e suggerisce che, nonostante una crescita continua, le condizioni socioeconomiche rimangono un fattore determinante nelle previsioni.

Il modello CBD (M6) fornisce ancora una volta le proiezioni più ottimistiche in tutte le categorie, sia per i maschi che per le femmine. Questo modello incorpora l'effetto coorte, il che implica che le generazioni più giovani potrebbero beneficiare di condizioni di vita e sanitarie migliori rispetto alle generazioni precedenti. Tuttavia, le differenze tra le classi pensionistiche rimangono evidenti. L'ottimismo che caratterizza il modello CBD(M6) è però ancora una volta accompagnato da una grande incertezza legata ai miglioramenti futuri delle condizioni di vita e sanitarie, soprattutto per le classi a reddito più basso.

Infine, il modello APC mostra una crescita più contenuta dell'aspettativa di vita rispetto al CBD, con una particolare tendenza alla stabilizzazione dopo il 2025, soprattutto per gli individui con pensioni basse. Per i maschi con pensioni basse, l'aspettativa di vita si ferma intorno ai 20 anni, con una curva leggermente decrescente, mentre per quelli con pensioni più elevate si stabilizza intorno ai 25-26 anni. Le femmine presentano valori più alti, ma anche qui si osserva una stabilizzazione intorno ai 30 anni per le pensioni basse e fino a 32-33 anni per quelle più alte. L'ampiezza dell'intervallo di confidenza nel modello APC è particolarmente significativa, soprattutto per le pensioni più basse, suggerendo un'elevata incertezza.

4.4.2.4 Considerazioni finali per l'aspettativa di vita.

L'analisi complessiva delle proiezioni dell'aspettativa di vita per le varie combinazioni di sesso-categoria pensionistica, sesso-gestione previdenziale, e sesso-classe d'importo della

pensione ha evidenziato importanti differenze legate alle condizioni socioeconomiche, al settore lavorativo e all'ammontare della pensione. In generale, le donne mostrano un'aspettativa di vita più alta rispetto agli uomini in tutte le combinazioni, riflettendo una tendenza consolidata di maggiore longevità femminile. Tuttavia, all'interno di ciascun gruppo, emergono differenze significative che riflettono il ruolo fondamentale delle condizioni lavorative, del reddito e delle dinamiche previdenziali.

Per quanto riguarda i modelli utilizzati, i risultati del modello CBD (M6), che incorpora l'effetto coorte, forniscono le previsioni più ottimistiche, suggerendo che le generazioni future beneficeranno di condizioni di vita e sanitarie migliori rispetto alle coorti precedenti. Tuttavia, l'incertezza associata a queste proiezioni è elevata, soprattutto per le categorie socio-economicamente più vulnerabili. Al contrario, il modello APC mostra una tendenza alla stabilizzazione della crescita dell'aspettativa di vita, particolarmente per gli individui con pensioni più basse o appartenenti a gestioni previdenziali meno avvantaggiate. Infine, i modelli più conservativi come ARIMA e Lee-Carter forniscono proiezioni più stabili, ma non considerano i miglioramenti generazionali in modo così marcato.

In definitiva, le differenze osservate tra le varie combinazioni e modelli sottolineano l'importanza di un'analisi multidimensionale della mortalità e della longevità. Le politiche previdenziali e sanitarie dovrebbero tener conto non solo delle tendenze generali, ma anche delle disuguaglianze socio-economiche e delle peculiarità di ciascun gruppo, per garantire che i miglioramenti nelle condizioni di vita e sanitarie siano distribuiti in modo equo tra tutte le categorie di pensionati.

4.4.3 Proiezione dell'indice E-Dagger (e^\dagger).

Dopo aver analizzato l'andamento delle proiezioni dell'aspettativa di vita (e_x), è altrettanto importante valutare un altro indicatore chiave della disuguaglianza nella sopravvivenza: l'indice e-dagger (e^\dagger). Questo indice, che misura la dispersione della lunghezza della vita all'interno di una popolazione, fornisce informazioni fondamentali su quanto variabile sia la durata della vita tra gli individui di una stessa coorte. Mentre l'aspettativa di vita fornisce una misura media della longevità, l'indice e-dagger ci permette di valutare le differenze tra i membri di una popolazione, evidenziando il grado di disuguaglianza nella sopravvivenza.

Studiare l'e-dagger è cruciale perché una maggiore dispersione nella durata della vita rappresenta una vulnerabilità per le popolazioni, con implicazioni significative per la pianificazione previdenziale, sanitaria e sociale. Popolazioni con un'elevata disuguaglianza nella sopravvivenza richiedono politiche mirate per ridurre queste disparità e migliorare il benessere di gruppi più vulnerabili. Inoltre, la proiezione dell'indice e-dagger con modelli diversi, come ARIMA, Lee-Carter, CBD (M6) e APC, ci permette di comprendere meglio come le dinamiche temporali e generazionali influenzino la disuguaglianza nel tempo. Ogni modello ha un approccio differente nell'interpretare l'evoluzione delle disuguaglianze di sopravvivenza, e confrontare le proiezioni permette di identificare scenari futuri alternativi, che possono essere cruciali per la gestione del rischio di longevità e per affrontare le sfide di sostenibilità dei sistemi pensionistici.

In questo paragrafo verranno esaminate le proiezioni dell'indice e-dagger per diverse combinazioni di individui, suddivisi in base a sesso, categoria pensionistica, gestione previdenziale e classe d'importo della pensione. Come per l'aspettativa di vita, le proiezioni sono state effettuate utilizzando i quattro modelli distinti sopra menzionati. Ciascun modello contribuisce a una valutazione differenziata dell'andamento della disuguaglianza nella sopravvivenza, con differenti livelli di incertezza e sensibilità alle dinamiche temporali e generazionali.

Inoltre, è importante dire che l'aspettativa di vita e l'indice e-dagger, pur misurando aspetti differenti della sopravvivenza, sono strettamente collegati e offrono una visione complementare delle dinamiche di mortalità. In genere, un aumento dell'aspettativa di vita è accompagnato da una diminuzione dell'e-dagger, poiché il miglioramento delle condizioni di vita e sanitarie tende a ridurre le differenze nella mortalità tra gli individui, rendendo la sopravvivenza più omogenea. Tuttavia, quando l'aspettativa di vita aumenta, ma l'e-dagger non si riduce proporzionalmente (o addirittura aumenta), ciò indica che i benefici dell'aumento della longevità non sono distribuiti equamente all'interno della popolazione. In questo caso, alcune fasce di popolazione potrebbero vivere molto più a lungo rispetto ad altre, accentuando le disuguaglianze socioeconomiche o legate all'accesso alle cure sanitarie.

Analizzare l'aspettativa di vita insieme all'indice e-dagger permette di comprendere non solo quanto a lungo vivono gli individui, ma anche quanto equamente sono distribuiti questi anni di vita all'interno della popolazione. Confrontare le proiezioni dei due indicatori attraverso modelli differenti ci offre una visione più sfumata delle dinamiche future della sopravvivenza, evidenziando sia i potenziali miglioramenti sia le disuguaglianze che potrebbero persistere. Questo approccio aiuta a identificare le aree in cui interventi mirati possono ridurre le disparità nella longevità e migliorare il benessere complessivo della popolazione.

Successivamente, si procede all'analisi delle proiezioni dell'indice e-dagger per le diverse combinazioni.

4.4.3.1 Combinazione Sesso-Categoria pensionistica.

Proiezione indice E-Dagger: Maschi-Pensione di Vecchiaia

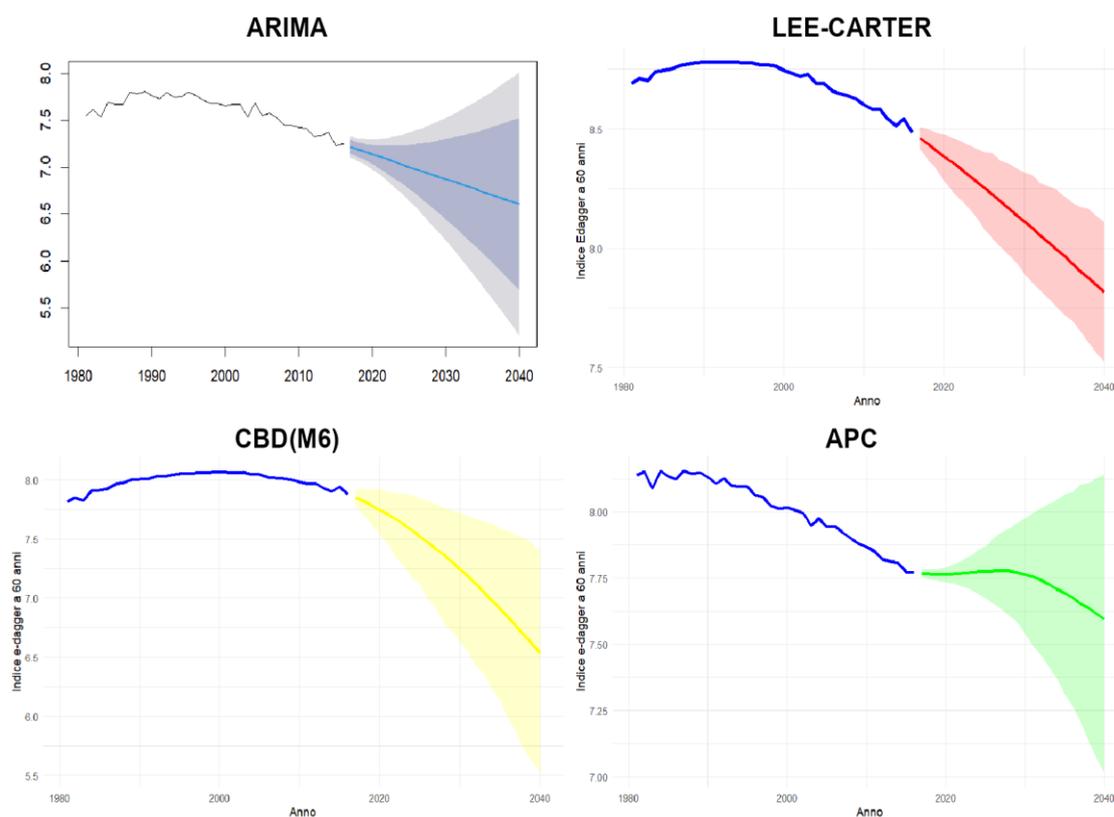


Figura 53: Proiezione indice E-Dagger: Maschi percettori di pensione di vecchiaia.

Proiezione indice E-Dagger: Maschi-Superstiti

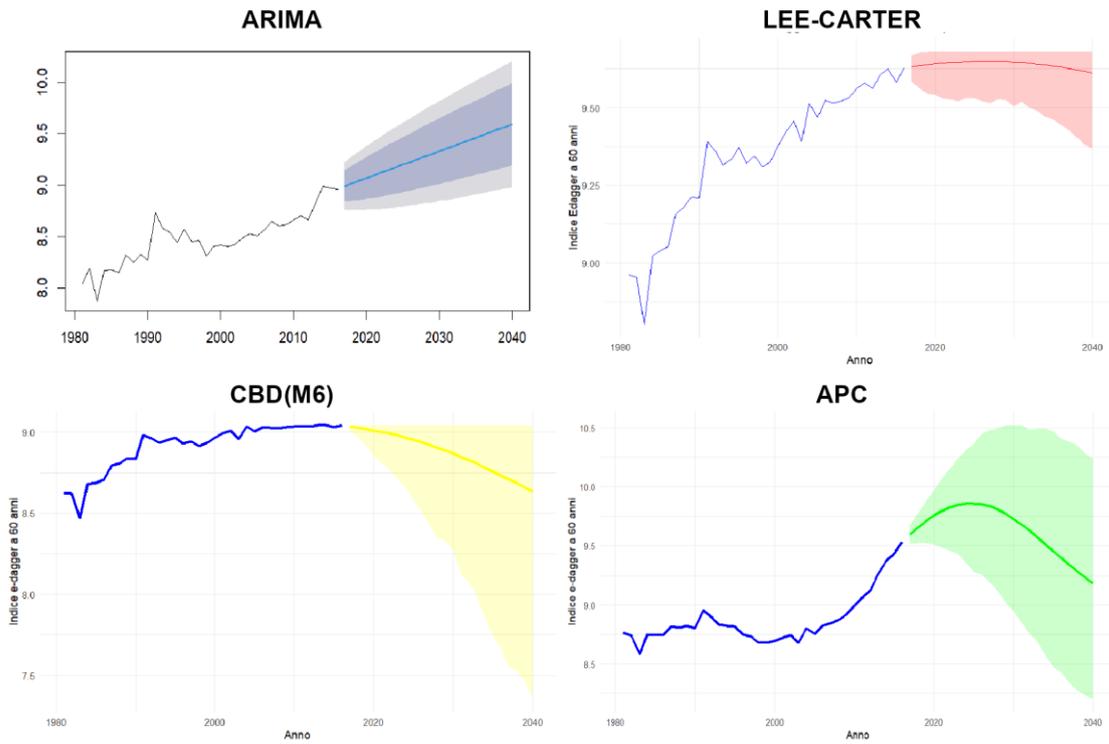


Figura 54: Proiezione indice E-Dagger: Maschi percettori di pensione ai superstiti.

Proiezione indice E-Dagger: Femmine-Pensione di Vecchiaia

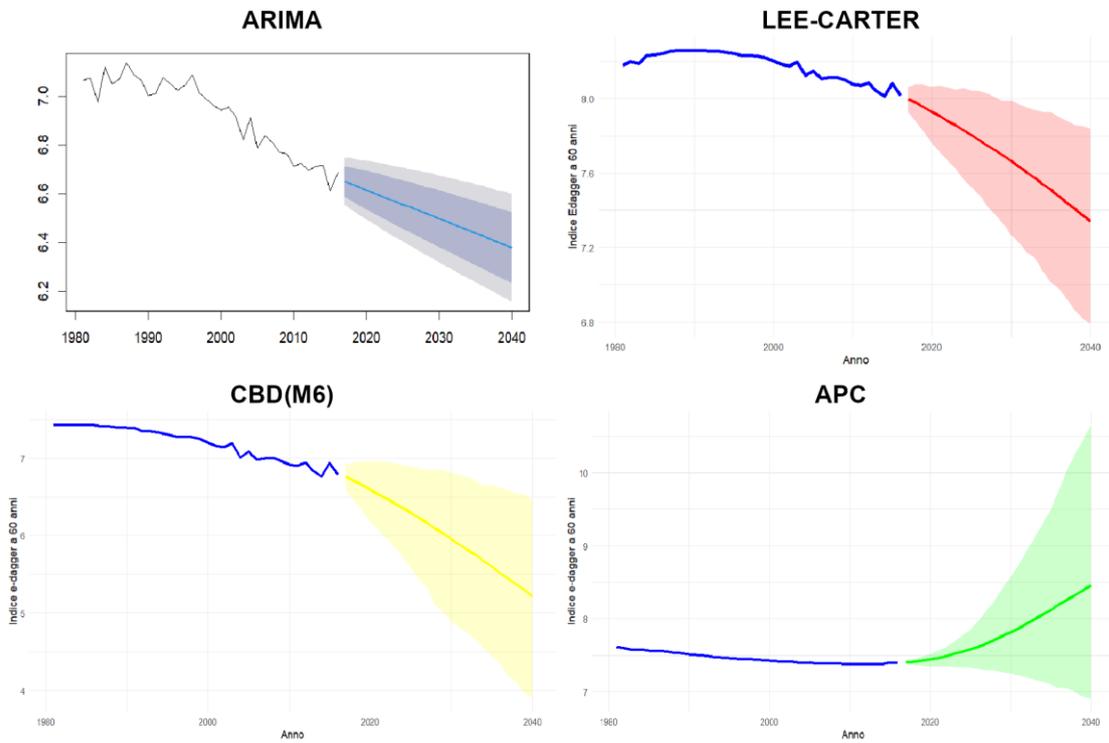


Figura 55: Proiezione indice E-Dagger: Femmine percettrici di pensione di vecchiaia

Proiezione indice E-Dagger: Femmine-Superstiti

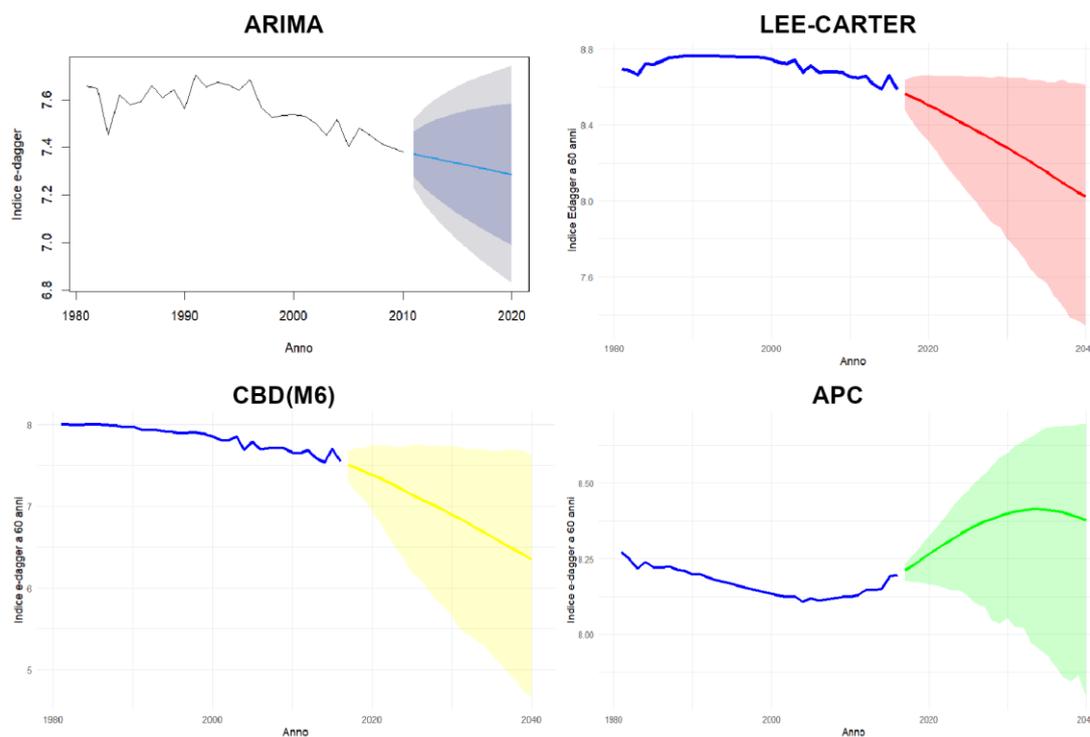


Figura 56: Proiezione indice E-Dagger: Femmine percettrici di pensione ai superstiti.

Le Figure 53, 54, 55 e 56 mostrano le proiezioni dell'indice e-dagger per quattro combinazioni di sesso e categoria pensionistica: rispettivamente Maschi – Vecchi (M-V), Maschi – Superstiti (M-S), Femmine – Vecchie (F-V) e Femmine – Superstiti (F-S). L'indice e-dagger, come ricordiamo, misura la dispersione nella durata della vita, evidenziando le disuguaglianze nella sopravvivenza all'interno di una popolazione. L'analisi delle proiezioni attraverso quattro modelli distinti (ARIMA, Lee-Carter, CBD (M6) e APC) rivela tendenze differenti, che dipendono non solo dalle condizioni sociali ed economiche delle categorie analizzate, ma anche dal funzionamento intrinseco di ciascun modello.

- Nel caso dei Maschi percettori di pensione di vecchiaia (M-V), tutti i modelli indicano una tendenza alla riduzione dell'indice e-dagger, suggerendo che le disuguaglianze nella sopravvivenza diminuiranno entro il 2040. Questo comportamento potrebbe riflettere un miglioramento complessivo delle condizioni di vita e sanitarie per questa categoria. Tuttavia, i modelli presentano delle forme diverse, dovute al modo in cui ciascuno gestisce le variabili temporali e demografiche. ARIMA, che si basa strettamente sui trend storici, mostra una riduzione graduale, ma con un ampio intervallo di confidenza, riflettendo un'alta incertezza per il lungo termine, tipica del suo approccio lineare che non considera cambiamenti strutturali. Al contrario, Lee-Carter prevede una riduzione più rapida

delle disuguaglianze, poiché tiene conto degli effetti temporali e cattura i miglioramenti sociali e sanitari che si sono verificati nel corso del tempo. Il modello CBD (M6), che incorpora l'effetto coorte, suggerisce un calo più pronunciato, ma con una maggiore incertezza, soprattutto verso il 2040, poiché cerca di proiettare i miglioramenti generazionali futuri, che sono difficili da prevedere con certezza. Infine, APC mostra una flessibilità maggiore, con un calo più dinamico ma con un intervallo di confidenza particolarmente ampio, riflettendo l'interazione tra età, periodo e coorte.

- Per i Maschi superstiti (M-S), le proiezioni sono più variegate. ARIMA si distingue dagli altri modelli per la sua previsione di un aumento dell'indice e-dagger, il che implica che le disuguaglianze nella durata della vita potrebbero crescere per questa categoria, suggerendo una persistenza delle disuguaglianze storiche. Questo riflette il fatto che, come già detto, ARIMA proietta le tendenze passate in modo lineare, senza considerare miglioramenti futuri strutturali. Al contrario, Lee-Carter e CBD (M6) prevedono una stabilizzazione o un leggero calo, segnalando che le disuguaglianze potrebbero ridursi leggermente nel tempo, ma senza la rapidità vista per i maschi percettori di pensione di vecchiaia. Il modello APC, con la sua natura complessa che incorpora sia l'età che il periodo e l'effetto coorte, mostra una fase di aumento iniziale seguita da una potenziale riduzione nel lungo termine, ma con un elevato grado di incertezza, rappresentato da intervalli di confidenza ampi, soprattutto dovuto alla complessità delle dinamiche di questa categoria, dove potrebbero esserci fluttuazioni legate a eventi generazionali o di periodo.
- Le proiezioni delle Femmine percettrici di pensione di vecchiaia (F-V) mostrano una tendenza simile a quella dei maschi vecchi, con una generale riduzione dell'e-dagger prevista da tutti i modelli, a eccezione del modello APC, che si distingue per la previsione di un aumento sostanziale dell'indice dopo il 2020. Questo aumento potrebbe riflettere nuove disuguaglianze emergenti tra le coorti più giovani di donne. Il modello Lee-Carter e il CBD (M6), al contrario, prevedono una riduzione più significativa, simile a quanto osservato per i maschi vecchi, suggerendo che il miglioramento delle condizioni sanitarie ridurrà le disuguaglianze nella sopravvivenza. Tuttavia, come nel caso dei maschi, il CBD (M6) mostra un'ampia incertezza dovuta al fatto che l'effetto coorte è difficile da prevedere con precisione. ARIMA, con la sua proiezione più lineare, indica una diminuzione graduale ma con meno dinamismo rispetto agli altri modelli, evidenziando il suo approccio conservativo e più semplice.
- Infine, le Femmine superstiti (F-S), presentano un quadro simile alle femmine di vecchiaia, ma con alcune variazioni importanti. Lee-Carter prevede una diminuzione piuttosto rapida dell'indice e-dagger, suggerendo che le disuguaglianze tra le superstiti si ridurranno significativamente entro il 2040. Il modello CBD (M6) segue una tendenza simile ma meno marcata, con una riduzione

più moderata. Invece, ancora una volta il modello APC prevede un aumento progressivo dell'indice, indicando che, anche per le femmine superstiti, potrebbero emergere nuove disuguaglianze generazionali, rendendo più variegata la distribuzione della durata della vita in questa categoria. ARIMA, ancora una volta, mostra una riduzione lineare dell'indice, ma con un'ampia incertezza, segnalando che le tendenze storiche potrebbero non essere affidabili per prevedere l'evoluzione delle disuguaglianze a lungo termine in questo gruppo.

A questo punto bisogna porsi una domanda molto importante, ovvero perché i modelli differiscono tra di loro? Ciò è dovuto al fatto che le differenze tra i modelli sono dovute principalmente al loro approccio specifico nell'affrontare le dinamiche di mortalità e disuguaglianza. ARIMA si basa su una proiezione lineare dei trend storici e tende a mantenere una certa inerzia, proiettando le tendenze passate senza considerare cambiamenti strutturali o dinamiche generazionali. Questo spiega perché ARIMA mostra tendenze più gradualmente e conservative, ma con ampi intervalli di confidenza, soprattutto a lungo termine. Lee-Carter, includendo l'effetto del periodo, cattura meglio i miglioramenti sociali e sanitari, soprattutto per le categorie che hanno beneficiato di tali cambiamenti, come i pensionati di vecchiaia. Tuttavia, non considera le differenze tra le coorti, il che lo rende meno adatto a catturare dinamiche generazionali. CBD (M6), invece, incorpora l'effetto coorte, il che lo rende più sensibile ai miglioramenti generazionali, ma con maggiore incertezza, poiché è difficile prevedere con precisione come evolveranno le condizioni di vita delle generazioni future. APC, infine, è il modello più complesso, poiché combina età, periodo e coorte, il che gli permette di catturare interazioni complesse tra questi fattori, ma allo stesso tempo genera una maggiore incertezza, soprattutto per categorie dove le dinamiche di disuguaglianza sono meno prevedibili, come ad esempio la categoria dei superstiti.

I motivi per i quali i tali dinamiche sono meno prevedibili per i superstiti, rispetto a coloro che percepiscono una pensione di vecchiaia, possono essere molteplici:

- In primo luogo, i superstiti rappresentano una categoria più eterogenea. A differenza dei pensionati di vecchiaia, che hanno una carriera lavorativa e contributiva alle spalle, i superstiti, spesso coniugi o familiari, possono avere condizioni economiche molto diverse. Alcuni superstiti possono aver beneficiato di un buon reddito dal coniuge, mentre altri potrebbero trovarsi in condizioni economiche più difficili, rendendo le proiezioni della loro durata di vita e disuguaglianza meno lineari e più difficili da prevedere.
- In secondo luogo, le pensioni dei superstiti sono calcolate sulla base della pensione del coniuge deceduto, quindi variano a seconda del reddito del defunto e del numero di beneficiari. Questa variabilità intrinseca introduce incertezza.
- Inoltre, i superstiti, e in particolare le vedove, tendono a essere più vulnerabili sia da un punto di vista sociale che economico. Dopo la perdita del coniuge, molti

superstiti vedono una riduzione significativa del reddito familiare, il che può influire negativamente sull'accesso alle cure sanitarie e sul benessere generale. Al contrario, i pensionati di vecchiaia godono generalmente di una maggiore stabilità economica, avendo accumulato risorse e benefici basati sulla propria carriera lavorativa.

- Un altro fattore è l'impatto psicologico della perdita del coniuge o del familiare, che può avere conseguenze negative sulla salute fisica e mentale del superstite. Questo impatto, essendo di natura propriamente soggettiva, non è facilmente quantificabile nei modelli di previsione, rendendo meno prevedibile l'evoluzione delle disuguaglianze tra i superstiti.

In sintesi, la maggiore eterogeneità, vulnerabilità socio-economica e l'impatto psicologico della perdita rendono le proiezioni delle disuguaglianze tra i superstiti molto più incerte rispetto ai pensionati di vecchiaia, che rappresentano una popolazione più stabile e omogenea dal punto di vista socio-economico.

4.4.3.2 Combinazione sesso – gestione previdenziale.

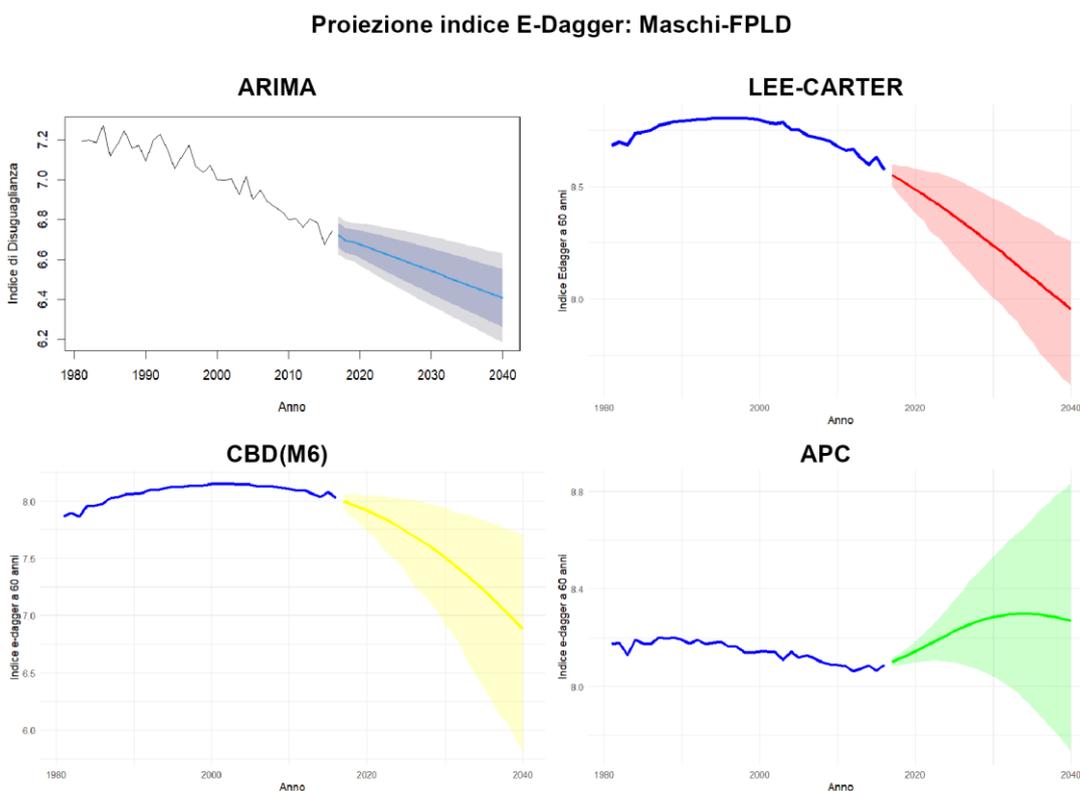


Figura 57: Proiezione indice E-Dagger: maschi appartenenti al Fondo Pensione Lavoratori Dipendenti.

Proiezione indice E-Dagger: Maschi-Coltivatori diretti, coloni e mezzadri

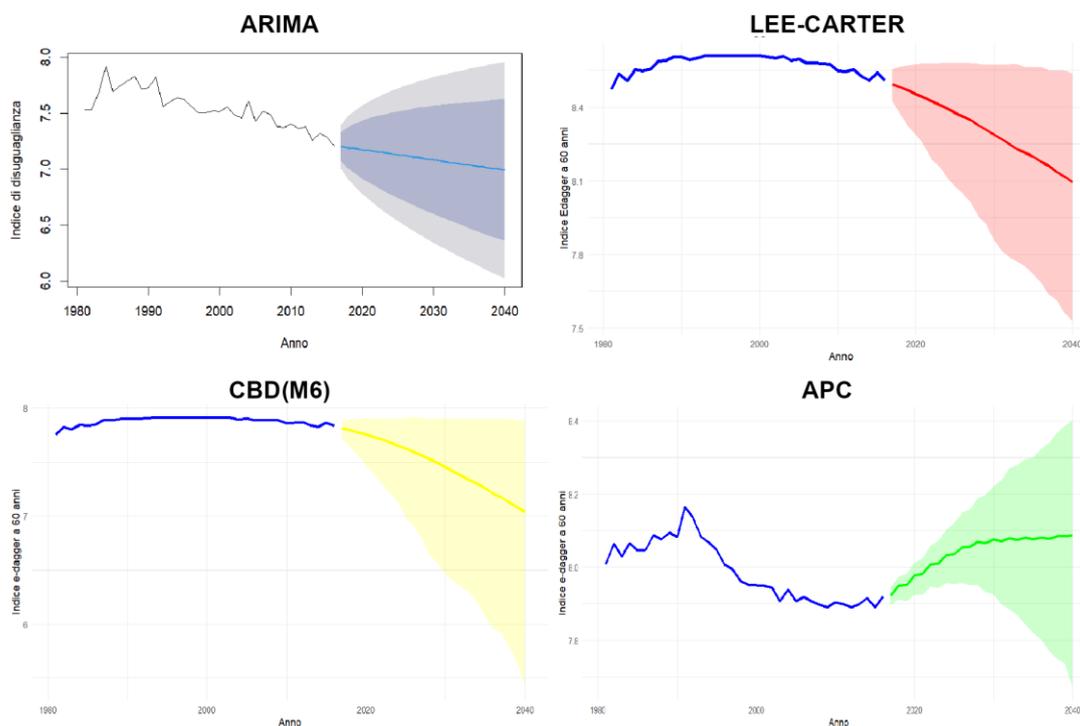


Figura 58: Proiezione indice E-Dagger: Maschi appartenenti ai Coltivatori diretti, Coloni e Mezzadri.

Proiezione indice E-Dagger: Maschi-Artigiani e Commercianti

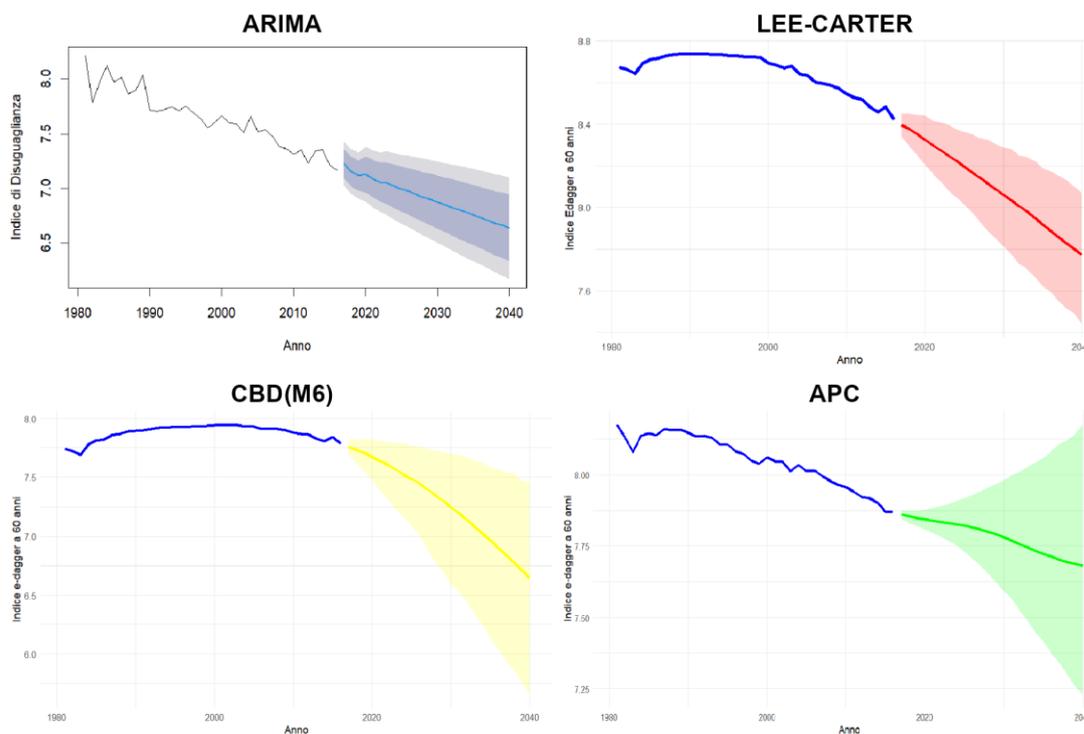


Figura 59: proiezione indice E-Dagger: Maschi appartenenti al gruppo degli Artigiani e Commercianti.

Le Figure 57, 58 e 59 mostrano i grafici delle proiezioni dell'indice e-dagger per i maschi appartenenti a tre diverse gestioni previdenziali. Tali grafici mostrano tendenze differenti in base ai modelli di previsione utilizzati e alle caratteristiche delle diverse gestioni.

- Nel caso dei maschi appartenenti al FPLD, i modelli indicano un generale declino dell'indice e-dagger, suggerendo una riduzione delle disuguaglianze nella sopravvivenza tra i lavoratori dipendenti nel lungo periodo. Il modello ARIMA prevede una riduzione piuttosto graduale dell'indice e-dagger, con un'ampia fascia di incertezza dopo il 2020. La tendenza alla riduzione graduale è plausibile per una categoria omogenea come quella dei lavoratori dipendenti, dove la stabilità dei benefici previdenziali contribuisce a una maggiore uniformità nella durata della vita.

Il modello Lee-Carter mostra una riduzione più netta e rapida. Questo modello, che incorpora l'effetto temporale, cattura bene i miglioramenti sociali e sanitari che si sono verificati nel tempo, portando a una riduzione più marcata delle disuguaglianze.

Il modello CBD (M6), invece, prevede un declino simile ma con un'ampia incertezza, soprattutto verso il 2040. Questo riflette l'importanza delle differenze generazionali, che il modello prende in considerazione. Nonostante una riduzione complessiva delle disuguaglianze, il modello suggerisce che le generazioni più giovani potrebbero sperimentare dinamiche diverse rispetto a quelle attuali, il che contribuisce all'aumento dell'incertezza nel lungo periodo.

Infine, anche il modello APC mostra una proiezione simile a quella vista con gli altri modelli, caratterizzata però da una maggiore incertezza molto più ampia verso la fine dell'intervallo di riferimento.

- Per i maschi appartenenti alla gestione dei coltivatori diretti, coloni e mezzadri, vengono evidenziate dinamiche più complesse. Questa categoria, infatti, rappresenta una popolazione più vulnerabile dal punto di vista economico e sociosanitario, il che si riflette nelle proiezioni dell'indice e-dagger.

Il modello ARIMA mostra una riduzione più graduale e meno marcata, accompagnata da un'ampia fascia di incertezza. Questo suggerisce che le disuguaglianze nella durata della vita potrebbero diminuire, ma in modo meno deciso rispetto ai lavoratori dipendenti.

Così come il modello ARIMA, anche il modello Lee-Carter prevede una riduzione delle disuguaglianze, ma in modo più moderato rispetto al caso dei maschi appartenenti al FPLD. Questo indica che, sebbene ci siano miglioramenti generali nella sopravvivenza, la popolazione dei coltivatori diretti, coloni e mezzadri continua a sperimentare maggiori disuguaglianze rispetto ai lavoratori dipendenti.

Il CBD (M6) segue una tendenza simile, con un calo previsto delle disuguaglianze, ma con una maggiore incertezza verso il 2040. Questo riflette la difficoltà di prevedere l'evoluzione delle coorti più giovani in una popolazione socio-economicamente più instabile. Anche se il modello suggerisce una riduzione, i futuri cambiamenti nelle condizioni economiche o sociali potrebbero influenzare significativamente queste previsioni.

Il modello APC, come ci si potrebbe aspettare per una popolazione più vulnerabile, mostra una curva più dinamica: dopo un periodo iniziale di stabilità, l'indice e-dagger aumenta leggermente prima di scendere di nuovo. Questo andamento riflette l'incertezza generata dall'interazione tra età, periodo e coorte, poiché i coltivatori diretti, coloni e mezzadri potrebbero risentire maggiormente delle fluttuazioni economiche e delle politiche sociali.

- Per i maschi appartenenti alla gestione degli artigiani e commercianti, i modelli presentano una tendenza simile a quella osservata per i maschi appartenenti al FPLD, ma con alcune differenze dovute alla variabilità delle condizioni economiche e previdenziali in questa categoria.

Il modello ARIMA, così come il modello Lee-Carter, prevedono una riduzione lenta e graduale dell'indice e-dagger, simile a quanto osservato per il FPLD.

Questo suggerisce che, nonostante le differenze economiche, anche gli artigiani e commercianti potrebbero beneficiare di miglioramenti uniformi nelle condizioni sanitarie e previdenziali.

Infine, il modello CBD (M6) e il modello APC prevedono una riduzione simile, ma con un margine di incertezza significativo. Questo è dovuto alla natura più instabile delle condizioni socio-economiche degli artigiani e commercianti, dove l'interazione tra età, periodo e coorte potrebbe portare a fluttuazioni significative nell'evoluzione della sopravvivenza.

Per i maschi appartenenti alle diverse gestioni previdenziali, tutti i modelli indicano una tendenza generale verso la riduzione delle disuguaglianze nella durata della vita, ma con ritmi e livelli di incertezza diversi. I lavoratori del FPLD mostrano una riduzione della proiezione dell'indicatore della Life-span Inequality, più marcata e prevedibile, grazie alla maggiore omogeneità della loro popolazione e alla stabilità dei benefici previdenziali.

Inoltre, si può affermare che il modello più ottimistico in questo caso è il modello Lee-Carter. Infatti, a differenza degli altri modelli, che presentano una maggiore incertezza nel lungo termine, il Lee-Carter mostra una riduzione consistente delle disuguaglianze con una fascia di incertezza più contenuta, rendendolo il modello che prevede il miglior scenario in termini di uniformità della sopravvivenza.

Successivamente si passa ad analizzare il caso delle femmine suddivise per gestione previdenziale.

Proiezione indice E-Dagger: Femmine-FPLD

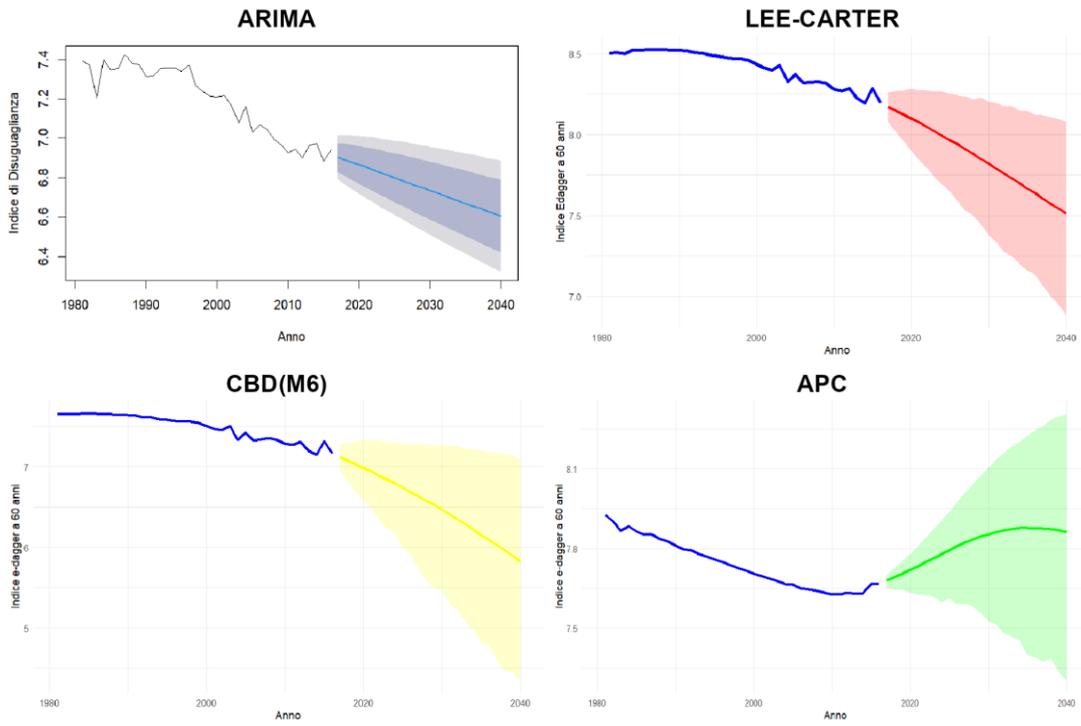


Figura 60: Proiezione indice E-dagger: Femmine appartenenti al Fondo Pensione Lavoratori Dipendenti.

Proiezione indice E-Dagger: Femmine-Coltivatori diretti, Coloni e Mezzadri

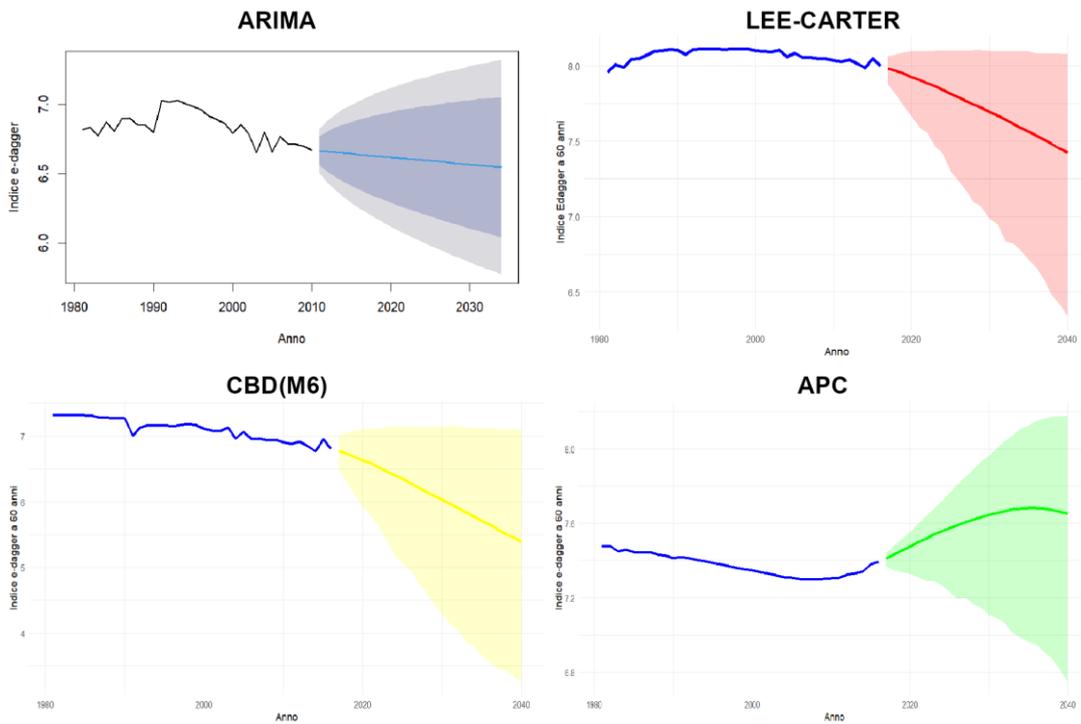


Figura 61: Proiezione indice E-Dagger: Femmine appartenenti alla gestione dei Coltivatori diretti, Coloni e Mezzadri.

Proiezione indice E-Dagger: Femmine-Artigiani e Commercianti

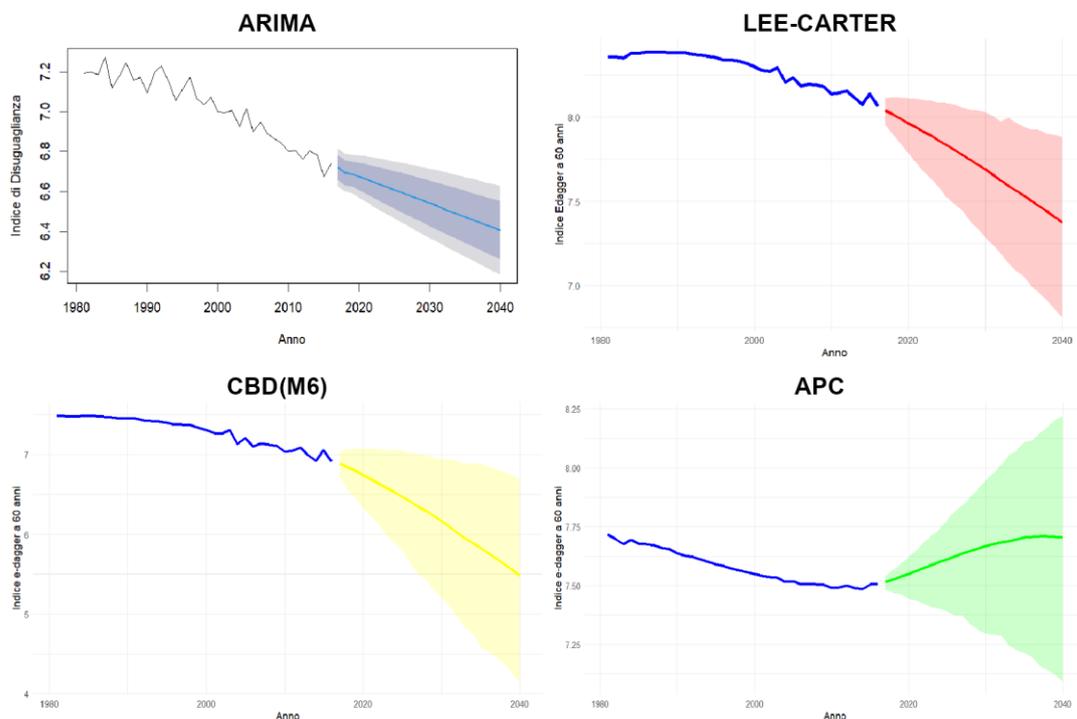


Figura 62: Proiezione indice E-Dagger: Femmine appartenenti alla gestione degli Artigiani e Commercianti.

Le proiezioni dell'indice e-dagger per le donne appartenenti alle tre diverse gestioni previdenziali mostrano una tendenza generale verso la riduzione delle disuguaglianze nella durata della vita, anche se con notevoli differenze a seconda della gestione e del modello utilizzato.

- Per le femmine appartenenti al FPLD, si osserva una riduzione dell'indice e-dagger più decisa e prevedibile rispetto alle altre gestioni. Il modello Lee-Carter mostra una forte diminuzione delle disuguaglianze già a partire dal 2020, riflettendo probabilmente l'uniformità con cui le donne in questa categoria hanno accesso ai benefici pensionistici e ai servizi sanitari. Questo modello cattura bene l'idea che, grazie ai miglioramenti socio-sanitari e previdenziali, le differenze nella sopravvivenza tra le donne in questa categoria si ridurranno più rapidamente. Anche il modello CBD (M6) conferma una tendenza simile, sebbene con una fascia di incertezza più ampia, soprattutto verso il 2040. Questo riflette la variabilità generazionale e l'incertezza su come le coorti future possano beneficiare di cambiamenti positivi. ARIMA, d'altra parte, prevede una riduzione graduale, con un andamento più lineare e una minore sensibilità ai cambiamenti strutturali, ma con una notevole incertezza nel lungo termine. Il modello APC aggiunge una dimensione interessante con una curva più dinamica, segnalando che i

cambiamenti generazionali potrebbero influenzare maggiormente le disuguaglianze, soprattutto dopo il 2020.

- La gestione dei coltivatori diretti, coloni e mezzadri evidenzia una maggiore incertezza rispetto al FPLD, con modelli che riflettono una popolazione più vulnerabile. Anche in questo caso, il Lee-Carter prevede una riduzione delle disuguaglianze, ma con un ritmo più lento rispetto alle donne del FPLD. Ciò suggerisce che, sebbene si stiano facendo progressi, le donne in questa gestione continueranno a sperimentare disuguaglianze più profonde. Il modello CBD (M6) segue una tendenza simile, ma con un'ampia fascia di incertezza, che riflette la vulnerabilità socio-economica della popolazione. ARIMA, come negli altri casi, mostra un calo graduale ma più incerto, poiché proietta tendenze storiche senza considerare pienamente le particolari dinamiche legate a età e periodo. Infine, il modello APC, invece, indica una certa stabilità iniziale dell'e-dagger prima di un successivo calo, suggerendo che le donne in questa gestione potrebbero sperimentare un lungo periodo di disuguaglianze persistenti prima di vedere miglioramenti.
- Per le femmine appartenenti alla gestione degli artigiani e commercianti, i modelli evidenziano un comportamento intermedio. Anche in questo caso i modelli Lee-Carter e CBD(M6) suggeriscono una riduzione rapida, con il secondo che presenta ancora una volta maggiore incertezza. Il modello ARIMA presenta una riduzione più lineare e il modello APC, ancora una volta suggerisce un andamento più dinamico, segnalando che fattori legati all'età e alla coorte potrebbero influenzare significativamente l'evoluzione delle disuguaglianze, soprattutto dopo il 2020.

Quindi, si può affermare che le proiezioni dell'indice e-dagger per maschi e femmine nelle diverse gestioni previdenziali mostrano un quadro complessivamente positivo, con una tendenza alla riduzione delle disuguaglianze nella sopravvivenza, anche se con dinamiche diverse tra i sessi e le gestioni. Per entrambe le categorie, il FPLD si distingue per una maggiore prevedibilità e stabilità nella riduzione delle disuguaglianze, grazie alla relativa omogeneità socio-economica della popolazione, con miglioramenti più marcati nelle femmine, probabilmente dovuti a una maggiore uniformità nei benefici pensionistici e nell'accesso ai servizi. D'altro canto, i coltivatori diretti, coloni e mezzadri, sia maschi che femmine, presentano una maggiore incertezza, dovuta alla vulnerabilità socio-economica e a condizioni di vita meno stabili. Le previsioni per gli artigiani e commercianti si collocano in una posizione intermedia, con una riduzione delle disuguaglianze prevista, ma con ampie fasce di incertezza, soprattutto nel lungo termine, dovute alle fluttuazioni economiche che caratterizzano questa categoria. In generale, i modelli Lee-Carter e CBD (M6) risultano i più ottimistici per entrambi i sessi, indicando che i miglioramenti in campo sanitario e previdenziale potrebbero portare a una maggiore uniformità nella durata della vita, mentre ARIMA e APC evidenziano una maggiore variabilità nelle

proiezioni, segnalando che gli effetti generazionali potrebbero influenzare l'evoluzione delle disuguaglianze in modo più complesso.

4.4.3.3 Combinazione sesso- classe d'importo della pensione.

Proiezione indice E-Dagger: Maschi- pensione inferiore a 645,05 euro

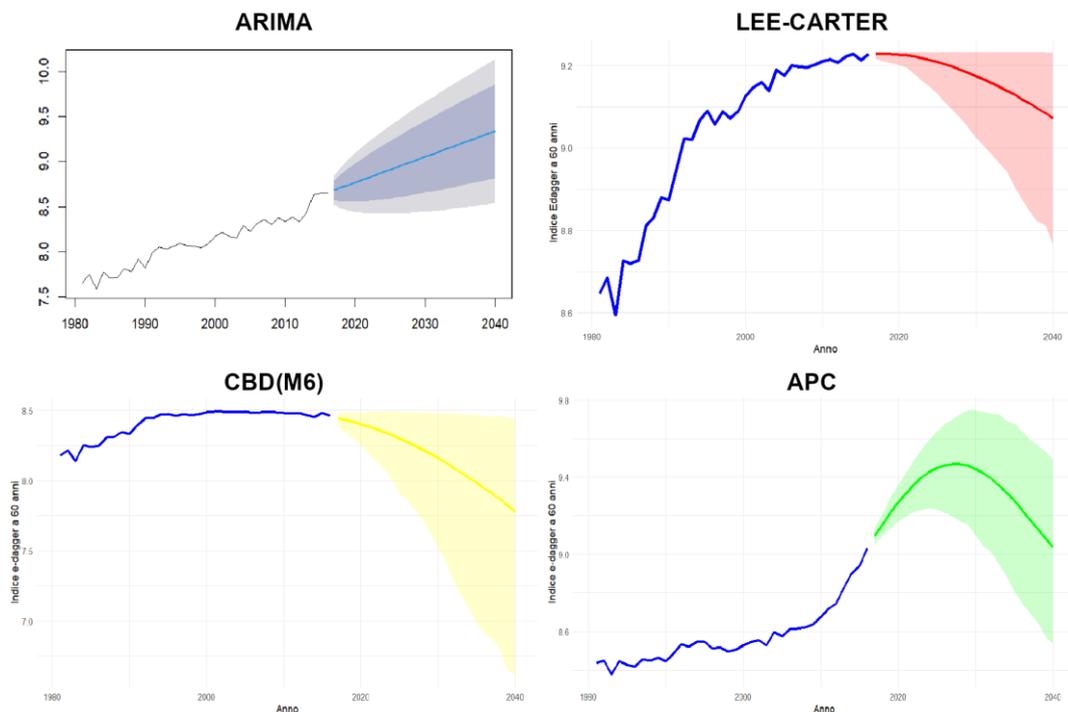


Figura 63: Proiezione indice E-Dagger: Maschi con pensione inferiore a 645,05 euro.

Proiezione indice E-Dagger: Maschi- pensione compresa tra 645,05 e 1290,10 euro

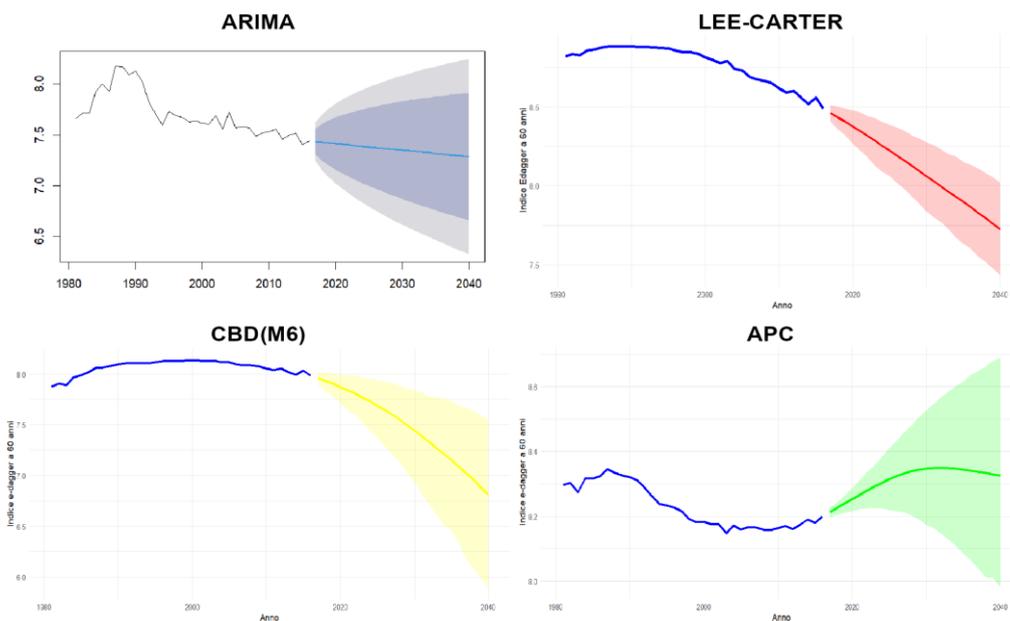


Figura 64: Proiezione indice E-Dagger: Maschi con pensione compresa tra 645,05 e 1290,10 euro.

Proiezione indice E-Dagger: Femmine- pensione inferiore a 645,05 euro

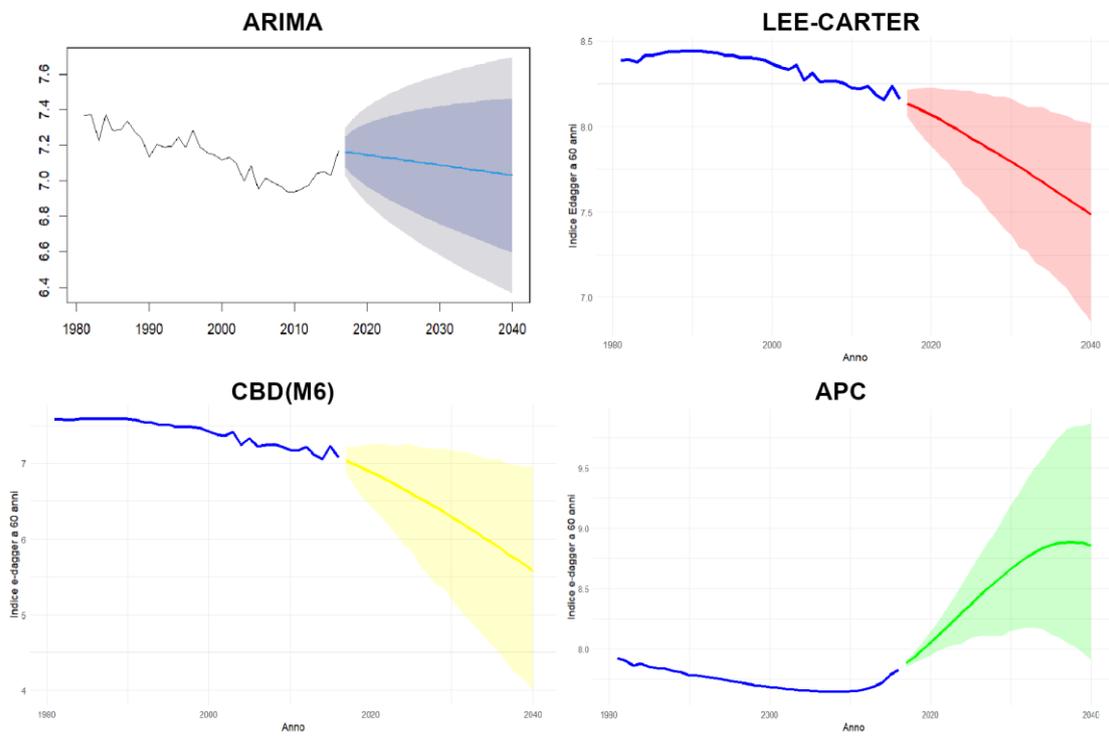


Figura 65: Proiezione indice E-Dagger: Femmine con pensione inferiore a 645,05 euro.

Proiezione indice E-Dagger: Femmine- pensione compresa tra 645,05 e 1290,10 euro

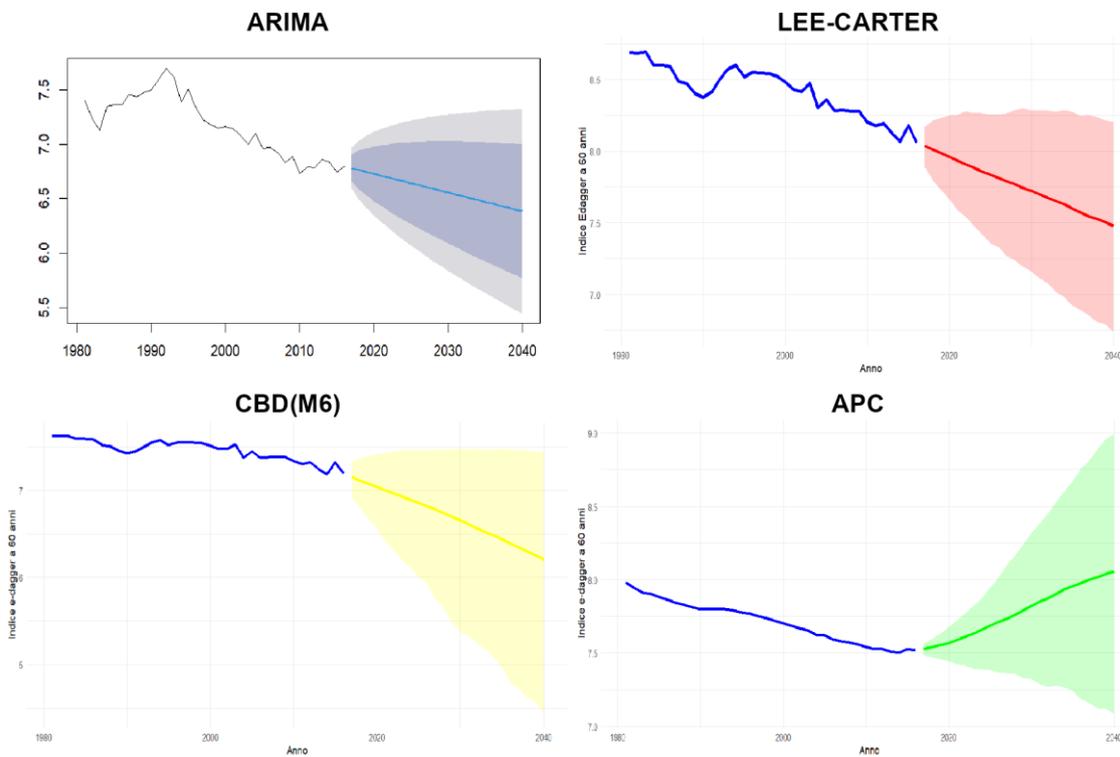


Figura 66: Proiezione indice E-Dagger: Femmine con pensione compresa tra 645,05 e 1290,10 euro.

Nella presente sezione, si analizzano le proiezioni dell'indice e-dagger in relazione alla combinazione sesso -classe d'importo della pensione (Figure 63, 64, 65 e 66), con lo scopo di valutare l'andamento delle disuguaglianze nella sopravvivenza tra i percettori di pensioni di importi differenti. La suddivisione per classe di importo consente di evidenziare come il livello di reddito influenzi la variabilità della durata della vita, considerando che pensioni più alte possono riflettere condizioni socioeconomiche più favorevoli e un miglior accesso a servizi sanitari.

- Nel caso delle pensioni inferiori a 645,05 euro, si nota che l'indice e-dagger tende a stabilizzarsi o addirittura a ridursi nel lungo termine in quasi tutti i modelli. ARIMA proietta un andamento gradualmente crescente per i maschi, con un'incertezza crescente nel lungo termine, mentre per le femmine si nota un lieve calo delle disuguaglianze, riflettendo forse un miglioramento delle condizioni socio-economiche di queste categorie nel tempo. Lee-Carter, invece, prevede un picco per i maschi, seguito da una riduzione marcata, indicando un futuro miglioramento significativo dell'omogeneità nella durata della vita per chi percepisce pensioni più basse. Anche per le femmine si osserva una riduzione, ma il calo è meno pronunciato, suggerendo che le disuguaglianze potrebbero persistere per un periodo più lungo. CBD (M6) mostra un andamento più stabile nel breve termine, ma una riduzione più evidente nel lungo periodo, con un'incertezza crescente verso il 2040, sia per i maschi che per le femmine. Il modello APC, d'altro canto, mostra per entrambi i sessi una fase di aumento iniziale delle disuguaglianze, seguita da una riduzione che però presenta un'ampia fascia di incertezza, suggerendo che le dinamiche di coorte e i fattori socio-economici futuri potrebbero influire in modo imprevedibile.
- Per quanto riguarda le pensioni comprese tra 645,05 e 1290,10 euro, le dinamiche cambiano leggermente. In generale, per questa fascia si osserva una riduzione delle disuguaglianze più evidente rispetto alla classe di pensioni più basse. ARIMA continua a mostrare una tendenza verso la riduzione per le femmine, mentre per i maschi l'incertezza è più pronunciata nel lungo periodo. Lee-Carter proietta una riduzione marcata, specialmente per i maschi, indicando che chi percepisce pensioni medie potrebbe beneficiare di condizioni di vita più omogenee. Il modello CBD (M6) segue un andamento simile, ma con una grande incertezza verso il futuro, segno che la variabilità tra le generazioni future rimane un fattore da considerare. Anche il modello APC per questa fascia evidenzia un comportamento dinamico, con una riduzione dell'e-dagger nel lungo termine, ma accompagnata da una fascia di incertezza ampia, soprattutto per le femmine, segnalando l'impatto dei cambiamenti di periodo e coorte nel tempo.

In riferimento alle proiezioni future, la minore disuguaglianza prevista per chi percepisce pensioni più alte si può spiegare considerando che le dinamiche sociali ed economiche a lungo termine tendono a consolidare i vantaggi di chi gode di una posizione più favorevole. Le proiezioni, in particolare nei modelli più ottimistici come Lee-Carter e CBD (M6), indicano che nel tempo chi percepisce pensioni più elevate continuerà a beneficiare di un migliore accesso a servizi sanitari e previdenziali, garantendo un miglioramento continuo nella qualità della vita e, di conseguenza, una maggiore uniformità nella durata della vita.

Le future generazioni di pensionati con redditi più alti saranno probabilmente esposte a migliori progressi in campo medico e tecnologico, che potranno essere distribuiti in modo più efficace tra coloro che dispongono di maggiori risorse. Ciò si traduce in proiezioni che vedono una riduzione più rapida delle disuguaglianze all'interno di queste categorie, poiché le innovazioni sanitarie e i miglioramenti nell'assistenza saranno accessibili in maniera più uniforme per chi ha maggiore capacità economica.

Le proiezioni a lungo termine riflettono quindi come il gap socio-economico contribuisca a ridurre le disuguaglianze per chi è già in una posizione privilegiata, mentre chi percepisce pensioni più basse potrebbe rimanere soggetto a maggiori variazioni nella durata della vita.

4.4.3.4 Considerazioni finali per l'Indice E-Dagger.

La valutazione delle proiezioni dell'indice e-dagger per le diverse combinazioni di sesso, categoria pensionistica, gestione previdenziale e classe di importo della pensione ha evidenziato tendenze comuni, ma con differenze significative tra i gruppi analizzati. In generale, si osserva una tendenza alla riduzione delle disuguaglianze nella durata della vita per la maggior parte delle categorie entro il 2040, con le proiezioni che indicano un miglioramento dovuto a fattori socio-economici e sanitari che, nel tempo, potrebbero ridurre le differenze nella sopravvivenza tra i gruppi. Tuttavia, l'entità e il ritmo di questa riduzione variano a seconda del modello utilizzato e della popolazione analizzata.

Le proiezioni più ottimistiche, come quelle del modello Lee-Carter e del CBD (M6), prevedono una diminuzione più rapida delle disuguaglianze, specialmente per le categorie più omogenee e con pensioni più elevate. D'altro canto, il modello APC, che considera le dinamiche di coorte, età e periodo, presenta una maggiore incertezza, soprattutto per le popolazioni più vulnerabili. Le proiezioni future indicano che questi gruppi potrebbero continuare a sperimentare disuguaglianze più marcate, riflettendo le loro condizioni socio-economiche instabili. Questi risultati hanno importanti implicazioni per il sistema previdenziale. Le proiezioni di una riduzione delle disuguaglianze nella durata della vita potrebbero portare a politiche più mirate per garantire una distribuzione più equa dei benefici pensionistici e dei servizi sociali, specialmente per i gruppi più vulnerabili. In particolare, la maggiore incertezza rilevata per i percettori di pensioni più

basse e per le categorie vulnerabili suggerisce la necessità di un potenziamento delle misure di protezione sociale per queste fasce, al fine di ridurre le disuguaglianze di lungo periodo. Il sistema previdenziale potrebbe utilizzare queste proiezioni per calibrare meglio i propri interventi.

4.4.4 Confronto dei Modelli Stocastici di Mortalità: Analisi dei Valori BIC per le Diverse Combinazioni.

COMBINAZIONI / MODELLO	BIC		
	LEE-CARTER	CBD(M6)	APC
M-V	19064.88	20560.45	19233.75
M-S	15472.86	15812.76	15488.15
F-V	17036.4	17225.58	18648.09
F-S	17861.25	16987.8	18564.56
M-A	19190.27	21063.89	20008.61
M-B	15048.29	15699.59	15146.08
M-C	14899.27	16307.49	15221.58
F-A	19718.98	18082.54	20136.01
F-B	15253.59	15796.61	15512.06
F-C	14552.11	15249.37	14937.84
M-1	18124.41	20179.12	19760.43
M-2	17210.58	18286.5	17690.07
F-1	20110.78	18628.68	20346.8
F-2	18034.16	16007.83	16945.79

Figura 67: Valori del BIC per le varie combinazioni.

L'analisi dei valori del BIC (Bayesian Information Criterion) per le varie combinazioni di sesso, categoria pensionistica, gestione previdenziale, nei tre modelli stocastici di mortalità considerati (Lee-Carter, CBD (M6) e APC), offre spunti interessanti e consente di valutare quale modello si adatta meglio alle diverse categorie di popolazione.

Per quanto riguarda la combinazione sesso – categoria pensionistica, emerge una tendenza chiara: il modello Lee-Carter tende a ottenere i risultati migliori, soprattutto per i maschi. In particolare, per i maschi con pensione di vecchiaia e per quelli con pensione ai superstiti (M-S), il Lee-Carter presenta valori del BIC più bassi rispetto agli altri modelli, suggerendo che riesce a catturare meglio le dinamiche della mortalità e sopravvivenza per questi gruppi. Il modello CBD (M6), che enfatizza l'effetto coorte, mostra risultati meno favorevoli, specie per i maschi con pensione di vecchiaia, dove il divario di BIC è significativo. Per le femmine, la situazione si presenta in maniera diversa. Nel caso delle femmine con pensione di vecchiaia, il Lee-Carter rimane il modello migliore, ma il

vantaggio rispetto al CBD è meno marcato rispetto ai maschi, suggerendo che l'effetto coorte può avere maggiore rilevanza per le donne. Sorprendentemente, per le femmine con pensione di superstite, il CBD (M6) supera il Lee-Carter, dimostrando che per questo gruppo specifico, caratterizzato da una maggiore vulnerabilità economica e sociale, l'effetto coorte è un fattore determinante.

Passando alla combinazione sesso-gestione previdenziale, per i maschi appartenenti al FPLD il modello Lee-Carter si conferma ancora una volta come il più adatto, con un BIC molto più basso rispetto agli altri modelli. Questa categoria, composta principalmente da lavoratori dipendenti, è probabilmente più omogenea, il che favorisce modelli come il Lee-Carter, che si concentra su effetti periodici senza considerare le dinamiche di coorte. Per i maschi appartenenti alla gestione dei coltivatori diretti, coloni e mezzadri e per quelli della gestione artigiani e commercianti, il Lee-Carter continua a essere il modello con il BIC più basso, anche se il vantaggio rispetto all'APC è meno marcato, suggerendo che per questi gruppi socio-economicamente meno stabili, i fattori legati al periodo e alla coorte potrebbero avere un ruolo più importante, anche se non abbastanza da superare il Lee-Carter. Per le femmine del FPLD, il CBD (M6) emerge come il modello più adatto, segnalando che per le donne lavoratrici dipendenti l'effetto coorte assume un'importanza rilevante. Per le femmine appartenenti alle gestioni dei coltivatori diretti e degli artigiani e commercianti, il Lee-Carter si dimostra ancora una volta il modello migliore, anche se il CBD (M6) si avvicina di più rispetto ai maschi nelle stesse categorie.

La combinazione sesso-classe d'importo mostra risultati coerenti con le precedenti analisi. Per i maschi con pensioni inferiori a 645,05 euro, il Lee-Carter ha il BIC più basso, ma con una differenza meno marcata rispetto all'APC, suggerendo che i fattori legati alla coorte o alle dinamiche temporali potrebbero essere rilevanti per questa classe di pensionati. Il CBD (M6), pur essendo più ottimistico nelle sue proiezioni, si dimostra meno efficace in termini di BIC per i maschi con redditi più bassi. Per i maschi con pensioni tra 645,05 e 1.290,10 euro, il Lee-Carter mantiene il suo primato, con differenze più marcate rispetto agli altri modelli. Per quanto riguarda le femmine con pensioni inferiori a 645,05 euro, il CBD (M6) ottiene il BIC più basso, suggerendo che l'effetto coorte gioca un ruolo fondamentale per le donne con pensioni basse, mentre per quelle con pensioni tra 645,05 e 1.290,10 euro, il CBD è nettamente il migliore, segnalando la sua capacità di cogliere le dinamiche generazionali anche per le classi di reddito medio-basse.

In sintesi, l'analisi del BIC rivela che il modello Lee-Carter è generalmente il più parsimonioso e adatto per la maggior parte delle combinazioni, soprattutto per i maschi e per le categorie socio-economicamente più stabili come i lavoratori dipendenti. Il CBD (M6) emerge come il migliore in alcuni casi specifici, in particolare per le femmine con pensioni basse o categorie più vulnerabili come i superstiti, dove l'effetto coorte assume maggiore importanza. Il modello APC, pur offrendo una buona spiegazione delle dinamiche di coorte e di periodo, tende a essere meno competitivo in termini di BIC

rispetto agli altri modelli, ma rimane utile per alcune categorie socio-economicamente più instabili o esposte a fluttuazioni generazionali.

4.4.5 Analisi delle Traiettorie nelle Proiezioni.

Le proiezioni dell'aspettativa di vita e dell'indice e-dagger, ottenute attraverso il modello Lee-Carter, vengono approfondite in questo paragrafo al fine di analizzare eventuali cambiamenti significativi e incroci nelle traiettorie durante il periodo di proiezione. Il modello Lee-Carter è stato scelto per questa analisi in quanto ha dimostrato le migliori prestazioni in termini di BIC nella maggior parte delle combinazioni considerate. Questa scelta permette di garantire una maggiore affidabilità nelle proiezioni, rendendo il modello un punto di riferimento per lo studio delle dinamiche future di sopravvivenza.

L'analisi si concentrerà sull'andamento delle traiettorie proiettate, con l'obiettivo di individuare eventuali incroci tra le diverse categorie di individui. Tali incroci, che non si manifestano nel periodo storico, potrebbero emergere durante la fase di proiezione, segnalando cambiamenti significativi nelle dinamiche di sopravvivenza. La loro identificazione è fondamentale per comprendere le possibili variazioni nelle disuguaglianze di longevità tra i gruppi esaminati, offrendo preziose indicazioni sull'evoluzione della mortalità futura.

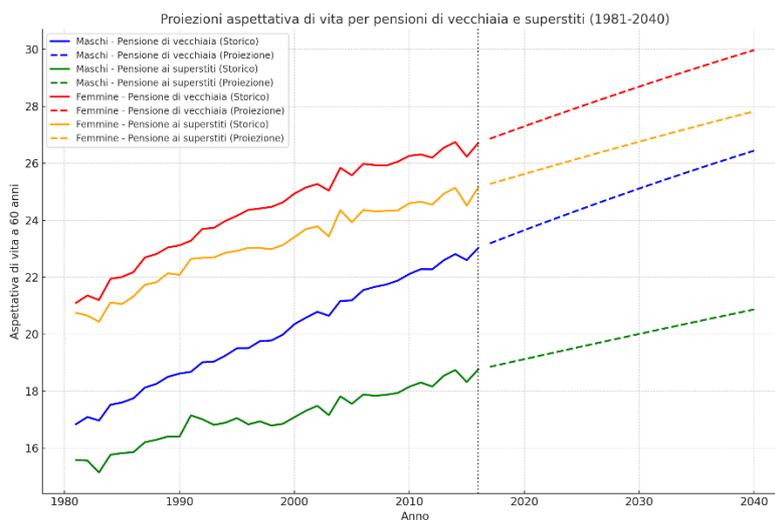


Figura 68: Andamento storico e previsto dell'aspettativa di vita: combinazione sesso-categoria pensionistica.

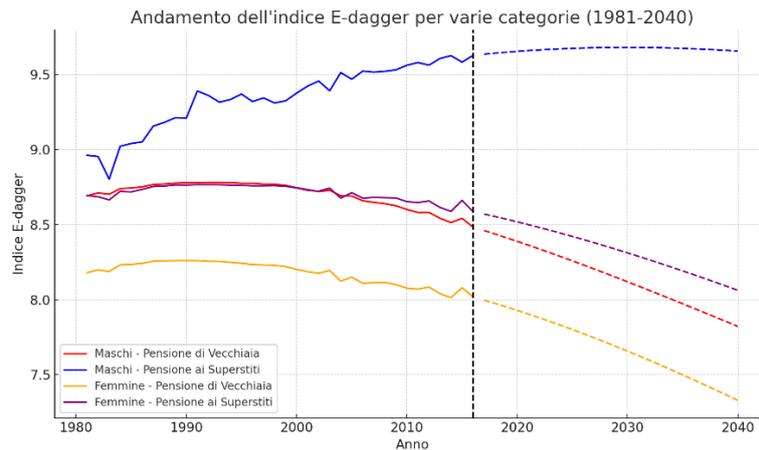


Figura 69: Andamento storico e previsto dell'indice e-dagger: combinazione sesso-categoria.

Analizzando i grafici delle proiezioni relative all'aspettativa di vita e all'indice E-Dagger per le pensioni di vecchiaia e di superstiti tra maschi e femmine, concentrandoci sul periodo di proiezione che va dal 2017 al 2040, emergono alcune dinamiche significative.

Per quanto riguarda l'aspettativa di vita, tutte le categorie considerate (maschi, femmine, pensioni di vecchiaia e superstiti) mostrano una traiettoria di crescita costante durante il periodo di proiezione. Le femmine con pensione di vecchiaia continuano a registrare i valori di aspettativa di vita più elevati, seguite dalle femmine con pensione di superstiti. I maschi con pensione di vecchiaia si trovano al terzo posto, seguiti dai maschi con pensione di superstiti, che rimangono costantemente, anche nel futuro, gli individui con un'aspettativa di vita minore. Per tali combinazioni, non si notano incroci tra le diverse traiettorie di proiezione, suggerendo che la gerarchia delle aspettative di vita osservata nella parte storica si mantiene anche nel periodo di proiezione, con le donne che continuano a godere di una maggiore longevità rispetto agli uomini, e i percettori di pensione di vecchiaia che vivono più a lungo rispetto ai superstiti.

Per quanto riguarda l'indice e-dagger, i grafici mostrano un andamento interessante. Anche in questo caso non si osservano incroci tra le proiezioni dei vari gruppi, ma ci sono delle tendenze divergenti. I maschi con pensione di superstiti partono da un livello molto più alto di disuguaglianza rispetto agli altri gruppi e continuano a mantenere questa posizione anche nel periodo proiettato fino al 2040, sebbene ci sia una stabilizzazione dell'indice verso la fine del periodo. Questo suggerisce che tra i maschi, la variabilità nella durata della vita rimarrà elevata anche in futuro. Le femmine, invece, vedono una significativa riduzione delle disuguaglianze nella sopravvivenza, con un calo costante dell'indice e-dagger nel corso della proiezione. Infine, i maschi con pensione di vecchiaia mostrano una tendenza simile, ma con una riduzione meno pronunciata rispetto alle donne.

In sintesi, durante il periodo di proiezione 2017-2040, non si osservano incroci tra le traiettorie dell'aspettativa di vita o dell'indice e-dagger. Quindi, anche nel periodo di proiezione è possibile osservare le stesse tendenze viste nel periodo storico.

Successivamente, ci si concentra sull'analisi delle proiezioni delle due grandezze di sopravvivenza per la combinazione sesso-gestione previdenziale. L'obiettivo è osservare l'evoluzione delle traiettorie di sopravvivenza e delle disuguaglianze nella durata della vita tra le diverse gestioni. Verranno esaminate le proiezioni a partire dal 2017 fino al 2040, con particolare attenzione a eventuali incroci nelle traiettorie che potrebbero indicare cambiamenti significativi nelle dinamiche di mortalità e sopravvivenza nel futuro.

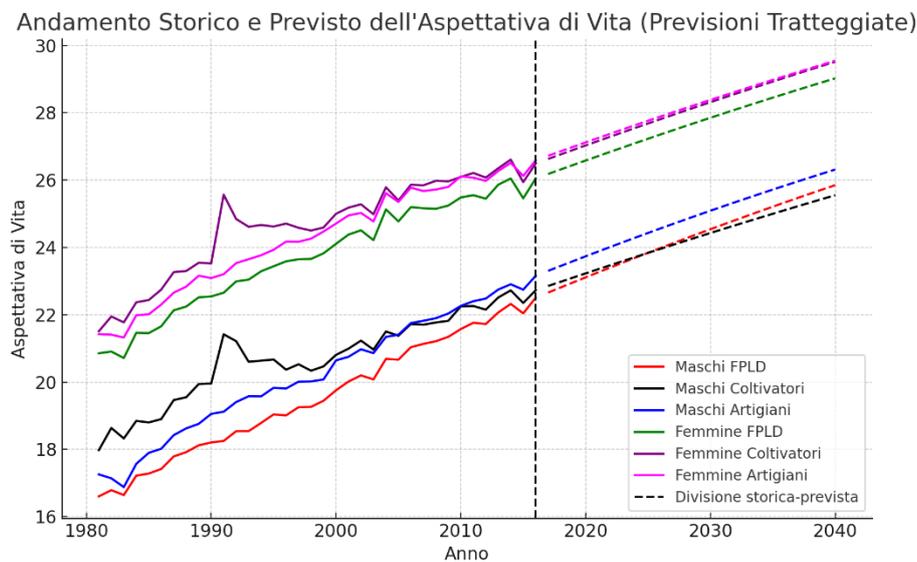


Figura 70: Andamento storico e previsto dell'aspettativa di vita: combinazione sesso-gestione previdenziale.

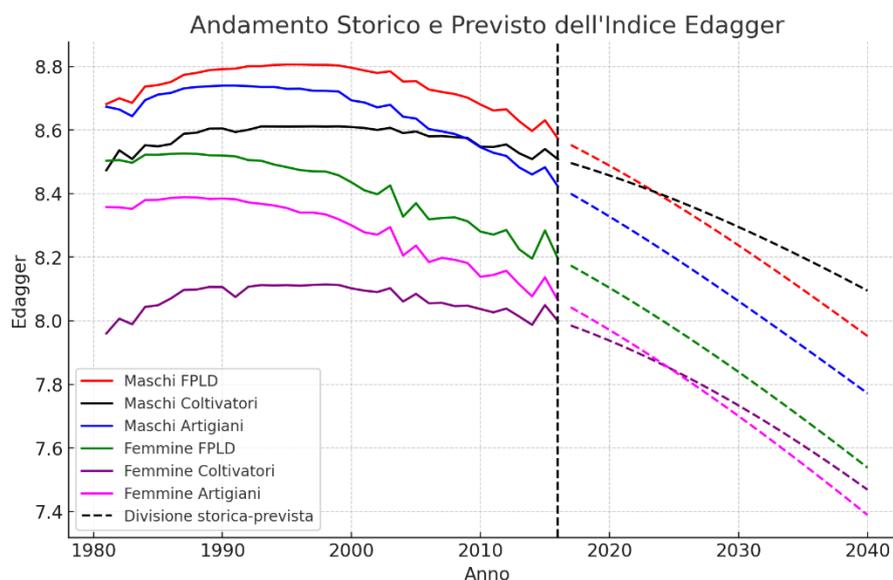


Figura 71: Andamento storico e previsto dell'indice e-dagger: combinazione sesso-gestione previdenziale.

Nel grafico delle proiezioni dell'aspettativa di vita per le combinazioni sesso-gestione previdenziale, tutte le categorie mostrano un trend di crescita costante fino al 2040. Si osservano alcune differenze tra i vari gruppi, soprattutto per quanto riguarda i maschi e le femmine di diverse gestioni. Anche nel periodo di proiezione è confermata la tendenza per il quale le donne presentano aspettative di vita più elevate.

Però, un elemento importante da descrivere è che durante il periodo di proiezione, ed in particolare alla metà degli anni '20, c'è un incrocio di traiettorie tra i maschi appartenenti al FPLD e i maschi coltivatori. Questo incrocio porta i maschi appartenenti alla gestione dei coltivatori diretti, coloni e mezzadri ad essere il gruppo che presenta un'aspettativa di vita inferiore alla fine del periodo considerato. Questo, evidenzia come tale categoria di individui abbia goduto di minori vantaggi derivanti dagli sviluppi socio-economici che ci sono stati nel tempo. Per quanto riguarda le donne invece, si osserva una piena convergenza tra coloro che appartengono al gruppo delle coltivatrici e coloro che invece sono artigiane e commercianti. Infatti, per tali categorie di individui, alla fine del periodo di osservazione è registrato un valore quasi uguale, poco superiore ai 29 anni.

Passando all'indice e-dagger, il grafico evidenzia una tendenza generale alla riduzione delle disuguaglianze nella durata della vita per tutte le categorie considerate. Però, anche in questo caso si osservano gli stessi incroci di traiettorie osservati per l'aspettativa di vita. In particolare, per le donne, che in generale presentano i valori di disuguaglianza più bassi, c'è un incrocio di traiettorie tra coloro che appartengono alla gestione relativa ai coltivatori diretti, coloni e mezzadri, e coloro che invece appartengono alla gestione degli artigiani e commercianti. Per questo le donne appartenenti a quest'ultima gestione alla fine del periodo registreranno un valore di disuguaglianza minore, evidenziando come esso sia un gruppo più omogeneo rispetto agli altri gruppi. Per i maschi invece, l'incrocio avviene tra la proiezione degli appartenenti alla gestione dei coltivatori con quella degli appartenenti alla gestione del FPLD. Quindi nel 2040, il gruppo che presenta una disuguaglianza nella durata della vita maggiore è il gruppo dei maschi appartenenti alla gestione dei coltivatori diretti, coloni e mezzadri.

In sostanza, da tale analisi si evidenzia che nel futuro le diverse combinazioni di individui godono tutte di miglioramenti, ma con velocità diverse.

Infine, si passa alla valutazione e al confronto delle proiezioni per la combinazione sesso-classe d'importo.

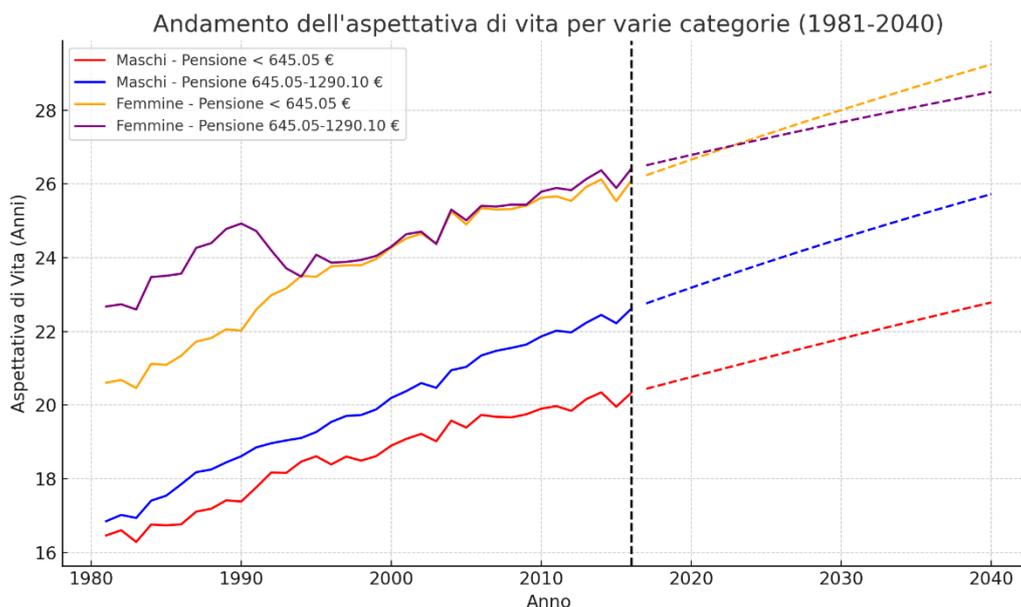


Figura 72: Andamento aspettativa di vita storica e prevista: combinazione sesso-classe d'importo.

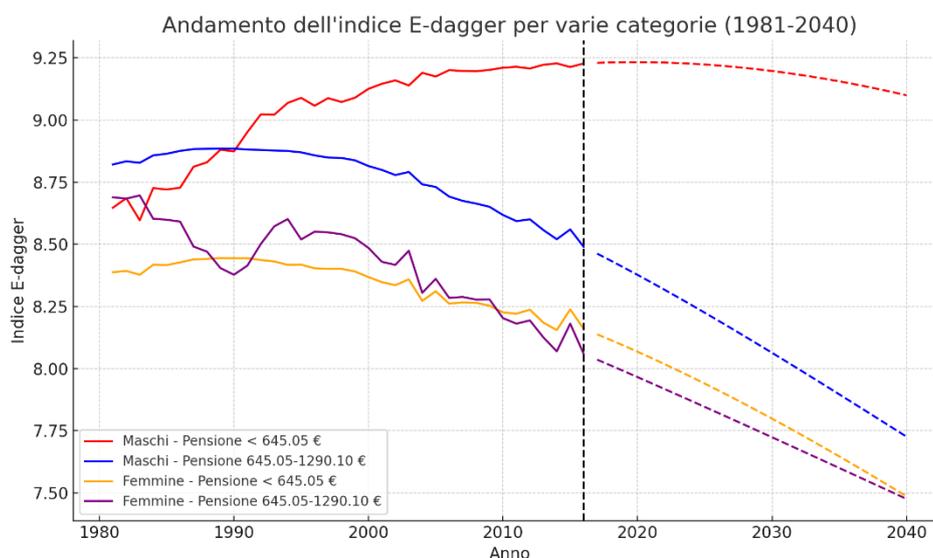


Figura 73: Andamento dell'indice e-dagger storico e previsto: combinazione sesso-classe d'importo

Nel grafico delle proiezioni dell'aspettativa di vita per la combinazione sesso-classe d'importo, si osservano alcune tendenze significative. La proiezione mostra chiaramente una crescita costante dell'aspettativa di vita in tutte le categorie considerate, con differenze evidenti tra le classi di pensione e il sesso degli individui. Gli individui con pensioni più alte (645,05-1.290,10 euro) tendono a mantenere una maggiore aspettativa di vita rispetto a quelli con pensioni inferiori a 645,05 euro. Le femmine, indipendentemente dalla classe d'importo, continuano a mostrare un'aspettativa di vita superiore rispetto ai maschi. In particolare, inizialmente le femmine con pensioni più alte hanno il livello più elevato di aspettativa di vita tra tutte le categorie, seguite dalle femmine con pensioni più basse.

Tuttavia, durante la fase di proiezione avviene un incrocio di traiettorie che porta le donne con pensione inferiore a 645,05 euro a godere di un'aspettativa di vita maggiore.

Per quanto riguarda i maschi, coloro che percepiscono pensioni più elevate mantengono un'aspettativa di vita maggiore rispetto ai maschi con pensioni più basse. Le due traiettorie mostrano un andamento parallelo durante tutto il periodo di proiezione e non si incrociano, anzi si allontanano, confermando che i maschi con redditi più alti continueranno a godere di una maggiore longevità rispetto a quelli con pensioni più basse.

Quindi mentre per gli uomini l'assenza di incroci tra le curve proiettate suggerisce che le gerarchie di longevità osservate in passato nella parte storica rimangono costanti, per le donne non è così. Infatti, per le femmine c'è un'inversione del trend che porta le donne con pensione inferiore a godere di un aumento maggiore dell'aspettativa di vita entro il 2040.

Il grafico relativo all'indice e-dagger per la combinazione sesso-classe d'importo evidenzia le tendenze nelle disuguaglianze di sopravvivenza fino al 2040.

Osservando le proiezioni, è evidente una riduzione generale dell'indice e-dagger in tutte le categorie a partire dal 2020, indicando che le disuguaglianze nella sopravvivenza tenderanno a diminuire nel lungo termine. Tuttavia, le traiettorie di questa riduzione variano tra i diversi gruppi.

I maschi con pensioni inferiori a 645,05 euro presentano l'indice e-dagger più elevato per tutto il periodo storico e anche nelle proiezioni future. Questo suggerisce che i maschi con pensioni basse continueranno a sperimentare le disuguaglianze più elevate in termini di durata della vita. Nonostante una lieve riduzione proiettata dopo il 2020, la loro disuguaglianza nella sopravvivenza rimane più alta rispetto agli altri gruppi. Mentre, per i maschi con pensioni più elevate (645,05-1.290,10 euro), l'indice e-dagger si abbassa più rapidamente rispetto a quello dei maschi con pensioni più basse.

Invece le femmine con pensioni più basse mostrano livelli di disuguaglianza inferiori rispetto ai maschi. Anche qui si osserva una riduzione dell'indice e-dagger, che sembra seguire una traiettoria costante di diminuzione nel periodo di proiezione, fino a raggiungere valori molto più bassi entro il 2040. Questo indica che le disuguaglianze nella sopravvivenza per le donne con pensioni più basse tenderanno a ridursi progressivamente. Infine, le femmine con pensioni più elevate mostrano un andamento simile, con l'indice e-dagger che scende ulteriormente nelle proiezioni. Le femmine con pensioni più alte sembrano beneficiare di una riduzione delle disuguaglianze di sopravvivenza più accentuata rispetto agli altri gruppi. Entro il 2040, l'indice e-dagger per questo gruppo si avvicina ai valori più bassi tra tutte le categorie.

In generale, non si osservano incroci significativi tra le proiezioni delle diverse categorie, suggerendo che le disuguaglianze di sopravvivenza, pur riducendosi, rimarranno differenziate in base al sesso e al livello della pensione. I gruppi con pensioni più elevate

continueranno a mostrare disuguaglianze minori rispetto a quelli con pensioni più basse, e le donne, in entrambe le fasce di reddito, sperimenteranno disuguaglianze inferiori rispetto agli uomini.

Quindi, le proiezioni mostrano differenze significative tra le combinazioni, evidenziando la necessità di politiche mirate per ridurre le disuguaglianze in longevità e sopravvivenza

CONCLUSIONE.

La tesi ha esplorato le dinamiche di mortalità e sopravvivenza dei percettori di pensione dell'INPS, attraverso un'analisi approfondita articolata in tre fasi principali: le statistiche descrittive, l'analisi storica delle grandezze di sopravvivenza, cioè l'aspettativa di vita e l'indice e-dagger, dal 1981 al 2016, e infine la proiezione di queste grandezze fino al 2040. Ogni fase ha contribuito a analizzare l'evoluzione della longevità e delle disuguaglianze nella durata della vita, ponendo l'accento su differenze significative tra gruppi socio-economici, di genere, di gestione previdenziale e di classe d'importo.

La **prima** fase ha riguardato le statistiche descrittive preliminari che hanno offerto una visione chiara delle caratteristiche socio-demografiche della popolazione pensionata nel periodo 1981 – 2016. Queste statistiche hanno rilevato differenze di genere significative nella distribuzione dei decessi. In particolare, il 57% dei decessi ha riguardato le donne, mentre il restante 43% ha coinvolto gli uomini. Questo dato riflette la maggiore longevità femminile, ovvero una tendenza ben documentata nella letteratura demografica. La maggiore aspettativa di vita delle donne comporta che esse rappresentino una porzione più ampia della popolazione anziana e, conseguentemente, dei decessi.

Un altro risultato chiave è emerso dall'analisi della tipologia di pensione. Tra gli uomini deceduti, l'81% percepiva una pensione di vecchiaia, mentre solo il 19% riceveva una pensione di superstite. Al contrario, tra le donne decedute, il 57% riceveva una pensione di superstite, riflettendo una realtà sociale in cui molte donne, specialmente tra le generazioni più anziane, avevano meno opportunità di carriera e si affidavano alle pensioni dei coniugi. Questo dato evidenzia una forte disparità di genere nelle carriere lavorative e nei contributi pensionistici, legata anche a ruoli tradizionali nel contesto familiare.

Inoltre, l'analisi del tasso di mortalità dal 1981 al 2016 ha mostrato un trend decrescente fino al 2005, attribuibile ai miglioramenti nelle condizioni di vita e nell'assistenza sanitaria. Le differenze di genere sono evidenti e significative anche nei tassi di mortalità: gli uomini hanno registrato tassi di mortalità più elevati rispetto alle donne per gran parte del periodo analizzato, riflettendo un maggiore coinvolgimento in attività lavorative fisicamente usuranti e una maggiore esposizione ai rischi. Tuttavia, a partire dal 2011, il tasso di mortalità femminile ha superato quello maschile, probabilmente a causa di un invecchiamento più marcato della popolazione femminile e di cambiamenti negli stili di vita. Questo sorpasso ha confermato che le dinamiche di mortalità tra i due sessi sono complesse e influenzate da numerosi fattori, sia demografici che socio-economici.

Successivamente, anche le classi d'importo delle pensioni hanno rivelato forti disparità di genere. Il 78% delle donne percepiva una pensione inferiore a 645,05 euro, mentre tra gli uomini solo il 47% apparteneva a questa classe d'importo. Inoltre, mentre una percentuale significativa di uomini (36%) riceveva pensioni comprese tra 645,05 e 1290,10 euro, solo il 21% delle donne apparteneva a questa fascia. Questo squilibrio è probabilmente il risultato

di carriere lavorative femminili caratterizzate da interruzioni, minori retribuzioni e un accesso limitato a posizioni lavorative ben remunerate. In aggiunta, tra le pensioni più elevate, la presenza maschile è nettamente superiore, con un 15% di uomini che riceve pensioni superiori a 1290,10 euro, rispetto a una quota quasi nulla tra le donne.

Un altro aspetto cruciale emerso dall'analisi descrittiva è la distribuzione delle gestioni previdenziali. Il Fondo Pensioni Lavoratori Dipendenti (FPLD) rappresenta la gestione principale per entrambi i sessi, ma con una prevalenza più marcata tra le donne. Questo dato riflette il fatto che le donne, soprattutto negli ultimi decenni, hanno partecipato principalmente al lavoro dipendente, mentre gli uomini hanno una rappresentanza leggermente maggiore nelle gestioni relative ai coltivatori diretti e agli artigiani.

La **seconda** fase dell'analisi ha studiato l'andamento della Life Expectancy (aspettativa di vita) e dell'indice E-Dagger tra il 1981 e il 2016, concentrandosi su diversi fattori di differenziazione come sesso, categoria pensionistica, gestione previdenziale e classe d'importo.

Per la combinazione sesso-categoria pensionistica l'analisi ha rivelato un generale aumento dell'aspettativa di vita a 60 anni per tutte le combinazioni considerate, sia per gli uomini che per le donne, con una tendenza marcata di miglioramento delle condizioni sanitarie e di vita. Le donne, come atteso, hanno costantemente registrato un'aspettativa di vita superiore rispetto agli uomini, sia tra i percettori di pensione di vecchiaia che di superstiti. Tuttavia, i percettori di pensione di vecchiaia, sia maschi che femmine, hanno mostrato una longevità maggiore rispetto ai percettori di pensione ai superstiti, probabilmente grazie a una storia lavorativa più stabile e condizioni socioeconomiche più favorevoli.

Per quanto riguarda l'indice e-dagger, che misura la disuguaglianza nella durata della vita, sono emerse importanti differenze di genere e di categoria. Gli uomini percettori di pensione ai superstiti hanno mostrato il valore più alto di e-dagger. Al contrario, le donne con pensione di vecchiaia hanno registrato i valori più bassi di e-dagger, evidenziando una maggiore equità nella distribuzione della longevità all'interno di questo gruppo. In generale la tendenza ha indicato una riduzione delle disuguaglianze di mortalità per quasi tutte le categorie, con un progresso più accentuato per le donne.

Per la combinazione sesso- gestione previdenziale, l'analisi ha rivelato che i coltivatori diretti, coloni e mezzadri inizialmente avevano un'aspettativa di vita superiore rispetto ai lavoratori dipendenti e artigiani/commercianti. Tuttavia, a partire dagli anni 2000, gli artigiani e commercianti hanno superato gli altri gruppi in termini di longevità, soprattutto per gli uomini. Questo cambiamento potrebbe riflettere un miglioramento delle condizioni economiche e sanitarie per queste categorie di lavoratori autonomi.

L'indice e-dagger, invece, ha mostrato una tendenza alla riduzione delle disuguaglianze in tutte le gestioni previdenziali. Tuttavia, gli artigiani e i commercianti hanno presentato i

valori più alti dell'indice di disuguaglianza negli anni '80, segnalando inizialmente una forte disuguaglianza nella durata della vita all'interno di questo gruppo, che si è poi ridotta verso la fine del periodo analizzato.

Per la terza ed ultima combinazione, ovvero quella relativa a sesso - classe d'importo, l'analisi ha confermato che chi percepisce pensioni più elevate tende a vivere più a lungo, mentre chi riceve pensioni inferiori ha 645,05 euro ha un'aspettativa di vita più bassa. Questo dato riflette come il reddito pensionistico sia un fattore determinante nella longevità

Inoltre, l'indice della life-span inequality ha evidenziato che le disuguaglianze nella durata della vita sono più accentuate tra i percettori di pensioni più basse, soprattutto tra gli uomini. L'indice e-dagger ha mostrato un andamento crescente per questa categoria, indicando che, nonostante alcuni individui stiano vivendo più a lungo, esiste una crescente disparità nella sopravvivenza all'interno di questo gruppo.

La **terza** fase dell'analisi ha esplorato le proiezioni dell'aspettativa di vita e dell'indice e-dagger fino al 2040, utilizzando quattro modelli stocastici principali: ARIMA, Lee-Carter, CBD(M6) e APC (Age-Period-Cohort). Questi modelli hanno permesso di delineare l'evoluzione futura della longevità e delle disuguaglianze nella durata della vita tra i diversi gruppi di pensionati INPS, considerando variabili come sesso, categoria di pensione, gestione previdenziale e classe d'importo della pensione. I risultati di queste proiezioni forniscono indicazioni cruciali per la gestione del Longevity Risk e per lo sviluppo di politiche previdenziali a lungo termine.

Le proiezioni per la combinazione di sesso e categoria pensionistica (pensione di vecchiaia vs pensione ai superstiti) confermano un generale aumento dell'aspettativa di vita fino al 2040 per tutte le categorie. Le donne continueranno a vivere più a lungo rispetto agli uomini, come indicato dai modelli Lee-Carter e CBD (M6). Infatti, le femmine con pensione di vecchiaia mostrano i valori più elevati di Life Expectancy, con proiezioni che indicano un'aspettativa di vita post 60 anni superiore ai 30 anni entro il 2040.

D'altro canto, i maschi percettori di pensione ai superstiti mostrano un'aspettativa di vita più bassa rispetto agli altri gruppi, con una proiezione che, pur in crescita, rimane distante dalle proiezioni femminili. Questo riflette la vulnerabilità economica e sanitaria di questa categoria, composta da individui che spesso non hanno accumulato risorse sufficienti durante la vita lavorativa e che, quindi, potrebbero non beneficiare appieno dei miglioramenti sociali e sanitari.

In termini di disuguaglianza nella durata della vita, misurata dall'indice e-dagger, i modelli hanno evidenziato tendenze contrastanti. Per le donne con pensione di vecchiaia, le proiezioni mostrano una riduzione significativa delle disuguaglianze, con l'indice e-dagger previsto in forte discesa entro il 2040. Questo suggerisce che, non solo le donne

vivranno più a lungo, ma la distribuzione della longevità all'interno di questo gruppo sarà più equa, con una maggiore uniformità nella durata della vita. Al contrario per i maschi con pensione di superstita, l'indice e-dagger è previsto in crescita. Questo indica che, nonostante un moderato aumento dell'aspettativa di vita, le disuguaglianze all'interno di questo gruppo aumenteranno, con un piccolo gruppo di individui che vivrà molto più a lungo rispetto alla media, mentre altri moriranno prematuramente. Questo risultato segnala la persistente fragilità socioeconomica di questo gruppo, che riflette una condizione di vulnerabilità non solo economica ma anche sanitaria.

Successivamente, le proiezioni dell'aspettativa di vita per i maschi e le femmine appartenenti alle diverse gestioni previdenziali (FPLD, Coltivatori diretti, Artigiani e Commercianti) evidenziano tendenze importanti. In generale, tutti i modelli mostrano un aumento costante dell'aspettativa di vita fino al 2040, anche se i livelli assoluti e le traiettorie di crescita variano sia in funzione della gestione previdenziale, sia in base al modello considerato. Il modello ARIMA, ad esempio, presenta previsioni più conservative e lineari, con una crescita costante ma meno accentuata dell'aspettativa di vita. Le sue previsioni tendono a essere più stabili, con valori che si attestano tra i 27 e i 29 anni per i maschi e i 30-31 anni per le femmine, a seconda della gestione. Il modello Lee-Carter, che tiene conto di età e periodo, offre proiezioni leggermente più contenute rispetto ad ARIMA, ma con una crescente incertezza nelle previsioni a lungo termine, soprattutto per le categorie più vulnerabili. Il modello CBD (M6) si distingue per la sua maggiore sensibilità all'effetto coorte, che comporta previsioni decisamente più ottimistiche, soprattutto per le nuove generazioni di lavoratori. Tuttavia, questa visione ottimistica è accompagnata da una maggiore incertezza nelle proiezioni a lungo termine, soprattutto a causa dell'effetto coorte, che rende le stime più sensibili ai miglioramenti delle condizioni generazionali. Inoltre, il modello APC offre una visione più equilibrata, con un'aspettativa di vita che cresce in tutte le categorie, ma con una stabilizzazione a partire dal 2025. Infatti, il modello APC prevede valori di Life Expectancy che si stabilizzano tra i 29-30 anni per i maschi e i 30-31 anni per le femmine, con una riduzione della crescita nelle fasi finali del periodo di proiezione. Però, anche in questo caso, gli intervalli di confidenza si ampliano considerevolmente nel lungo termine, suggerendo che le previsioni a lungo termine sono soggette a un'elevata incertezza.

In termini di differenze tra le gestioni previdenziali, i lavoratori dipendenti (FPLD) hanno mostrato una crescita più stabile dell'aspettativa di vita, con una minore variabilità tra i modelli rispetto alle altre categorie. Gli artigiani e commercianti, invece, presentano le proiezioni più elevate, in particolare nel modello CBD, che suggerisce un notevole incremento della longevità per questa categoria, favorita dal miglioramento delle condizioni economiche e di lavoro.

Un altro risultato chiave delle proiezioni è l'incertezza crescente nel lungo termine, particolarmente evidente nei modelli più complessi come Lee-Carter, CBD (M6) e APC.

Questa incertezza è legata alle difficoltà intrinseche nel prevedere fattori esterni che influenzano la mortalità, come i progressi tecnologici, le politiche sanitarie e i cambiamenti socioeconomici. Tale incertezza è particolarmente rilevante per i coltivatori diretti, il cui futuro dipende da numerosi fattori legati all'evoluzione del settore agricolo e al miglioramento delle condizioni lavorative.

Invece, i risultati relativi alla proiezione dell'indice e-dagger per la combinazione sesso-gestione previdenziale evidenziano vari aspetti principali. In primis tutti i modelli utilizzati per la proiezione dell'indice e-dagger mostrano una tendenza generale verso la riduzione delle disuguaglianze nella durata della vita entro il 2040. Questo risultato è incoraggiante, poiché indica che, nonostante le differenze socioeconomiche tra le varie gestioni previdenziali, le disuguaglianze nella sopravvivenza potrebbero ridursi gradualmente nel tempo, contribuendo a una maggiore equità nella distribuzione della longevità. I lavoratori dipendenti (FPLD), in entrambe le categorie di genere, hanno mostrato una riduzione più marcata e prevedibile delle disuguaglianze, grazie alla maggiore omogeneità socioeconomica e alla stabilità dei benefici previdenziali. Mentre, i coltivatori diretti, coloni e mezzadri presentano dinamiche più complesse, con una maggiore incertezza nelle proiezioni. La condizione di instabilità economica e sanitaria di questa categoria rende le previsioni più soggette a variazioni, come evidenziato dai modelli più dinamici come APC, che prevede anche un aumento temporaneo delle disuguaglianze prima di un eventuale miglioramento. Le femmine, in tutte le gestioni previdenziali, mostrano una tendenza più decisa alla riduzione dell'indice e-dagger rispetto ai maschi, riflettendo una maggiore uniformità nella durata della vita tra le donne. Tra i modelli utilizzati, il Lee-Carter e il CBD (M6) si sono dimostrati i più ottimistici nel prevedere una riduzione delle disuguaglianze, in particolare per i lavoratori dipendenti e gli artigiani. Il modello Lee-Carter ha evidenziato una riduzione più rapida delle disuguaglianze, grazie alla sua capacità di catturare i miglioramenti sociali e sanitari nel tempo. Il modello CBD, d'altro canto, ha evidenziato un'importante incertezza per il lungo termine, legata all'effetto coorte, ma ha previsto comunque un miglioramento complessivo delle disuguaglianze.

In sintesi, le proiezioni dell'indice e-dagger indicano una tendenza generale alla riduzione delle disuguaglianze nella sopravvivenza entro il 2040, ma con ritmi differenti tra le gestioni previdenziali e tra i sessi.

In conclusione, i risultati principali relativi alle proiezioni dell'aspettativa di vita e dell'indice e-dagger per la combinazione sesso-classe d'importo delle pensioni mettono in luce diverse dinamiche.

Le proiezioni dell'aspettativa di vita indicano una differenza significativa tra i percettori di pensioni basse e quelli con pensioni più elevate. Le donne, come già osservato in altre combinazioni, mostrano un'aspettativa di vita superiore rispetto agli uomini in entrambe

le classi d'importo. Tuttavia, l'aumento previsto dell'aspettativa di vita è generalmente più contenuto per chi percepisce pensioni più basse, con differenze tra i modelli stocastici. Il CBD (M6) fornisce le proiezioni più ottimistiche, mentre l'APC mostra una tendenza alla stabilizzazione nel lungo termine, specialmente per i redditi più bassi.

Le proiezioni dell'indice e-dagger evidenziano una tendenza alla riduzione delle disuguaglianze nella durata della vita, con dinamiche differenti tra le fasce pensionistiche. Chi percepisce pensioni più basse mostra un maggiore grado di incertezza e instabilità nelle proiezioni, con modelli come l'APC che indicano un possibile aumento temporaneo delle disuguaglianze prima di una riduzione. Al contrario, chi percepisce pensioni più elevate tende a godere di una riduzione più marcata delle disuguaglianze, in particolare nelle previsioni più ottimistiche del modello Lee-Carter e CBD (M6). Queste differenze riflettono come il reddito pensionistico influenzi l'accesso a migliori condizioni di vita e sanitarie, contribuendo a un miglioramento della sopravvivenza e a una maggiore uniformità nella durata della vita per le fasce più privilegiate.

In sintesi, le proiezioni future confermano che il reddito pensionistico gioca un ruolo cruciale nel determinare la durata della vita e la sua distribuzione. Le disuguaglianze sono destinate a ridursi più rapidamente per chi percepisce pensioni elevate, mentre chi ha redditi più bassi continuerà a sperimentare maggiore variabilità e incertezza nella durata della vita.

In conclusione, la tesi ha anche sottolineato come l'utilizzo di diversi modelli stocastici consenta di ottenere prospettive multiple sulle dinamiche future della mortalità. Mentre modelli più ottimistici come il CBD (M6) prevedono un miglioramento più rapido delle condizioni di sopravvivenza, modelli più conservativi come l'APC e l'ARIMA offrono una visione più prudente, segnalando che l'incertezza e la variabilità socio-economica potrebbero avere un impatto significativo sulle previsioni future.

Questa analisi rappresenta un contributo significativo per la gestione del rischio di longevità nel sistema previdenziale italiano. Le proiezioni fornite possono servire come base per politiche previdenziali mirate a ridurre le disuguaglianze e garantire una maggiore equità nella distribuzione dei benefici pensionistici, supportando la sostenibilità del sistema a lungo termine. In particolare, le istituzioni previdenziali potrebbero sfruttare questi dati per promuovere politiche che migliorino l'accesso a servizi sanitari e sociali per i gruppi più vulnerabili, con l'obiettivo di ridurre le disuguaglianze di sopravvivenza e garantire che i miglioramenti nelle condizioni di vita siano distribuiti in modo più equo tra tutta la popolazione pensionata.

BIBLIOGRAFIA.

- 1) Aburto, José Manuel, and Alyson van Raalte. "Lifespan dispersion in times of life expectancy fluctuation: the case of Central and Eastern Europe." *Demography* 55.6 (2018): 2071-2096
- 2) Bisetti, Emilio, and Carlo A. Favero. "Measuring the impact of longevity risk on pension systems: The case of Italy." *North American Actuarial Journal* 18.1 (2014): 87-103.
- 3) Bjerre DS. TREE-BASED MACHINE LEARNING METHODS FOR MODELING AND FORECASTING MORTALITY. *ASTIN Bulletin*. 2022;52(3):765-787
- 4) Booth, Heather, et al. "Lee-Carter mortality forecasting: a multi-country comparison of variants and extensions." *Demographic research* 15 (2006): 289-310.
- 5) Booth, Heather, Leonie Tickle, and Len Smith. "Evaluation of the variants of the Lee-Carter method of forecasting mortality: a multi-country comparison." *New Zealand Population Review* 31.1 (2005): 13-34.
- 6) Cairns, Andrew JG, David Blake, and Kevin Dowd. "Modelling and management of mortality risk: a review." *Scandinavian Actuarial Journal* 2008.2-3 (2008): 79-113.
- 7) Cairns, Andrew JG, et al. "A quantitative comparison of stochastic mortality models using data from England and Wales and the United States." *North American Actuarial Journal* 13.1 (2009): 1-35.
- 8) Caselli G., Vallin J., Wunsch G., "Demografia: la Dinamica delle Popolazioni", Carocci, 2001.
- 9) Caselli, G. "Mortalità e sopravvivenza in Italia dall'Unità agli anni'30." *Popolazione, Società e ambiente. Temi di Demografia storica italiana (secc XVII-XIX)* (1990): 275-310.
- 10) Caselli, Graziella. "Enciclopedia delle scienze sociali (1996)".
- 11) Chan, Wai-Sum, Johnny Siu-Hang Li, and Jackie Li. "The CBD mortality indexes: modeling and applications." *North American Actuarial Journal* 18.1 (2014): 38-58.
- 12) COVIP | Commissione di vigilanza sui Fondi Pensione. "L'evoluzione del sistema pensionistico in Italia" (aprile 2022).
- 13) D. G. Lippolis. "Una breve introduzione al Machine Learning". 13-gennaio-2021
- 14) D'Amato Valeria, and Emilia Di Lorenzo. "IL FENOMENO DELLA LONGEVITA' ED IL RISCHIO DI MODELLO: ANALISI E MISURA." *XX Ciclo del Dottorato in Matematica per l'Analisi economica e la Finanza, Università degli Studi di Napoli Federico II* (2008).
- 15) De Waegenare, Anja, Bertrand Melenberg, and Ralph Stevens. "Longevity risk." *De Economist* 158 (2010): 151-192.
- 16) Deprez, Philippe, Pavel V. Shevchenko, and Mario V. Wüthrich. "Machine learning techniques for mortality modeling." *European Actuarial Journal* 7 (2017): 337-352.
- 17) Girosi, Federico, and Gary King. "Understanding the Lee-Carter mortality forecasting method." *Gking. Harvard. Edu* (2007).
- 18) Glossario OMS della Promozione della Salute World Health Organization, Ginevra (1998).
- 19) Goodfellow, Ian. "Deep learning." (2016).

- 20) Istat, Glossario. "Istat (2005): Glossario." (2005).
- 21) ISTAT. "L'evoluzione demografica in Italia dall'Unità a oggi." (2019).
- 22) Levantesi, Susanna, and Massimiliano Menziotti. "Allungamento della vita media e rischio assicurativo." *Allungamento della vita media e rischio assicurativo*. Vol. 2. Edizioni Scientifiche Italiane, 2016. 1-92.
- 23) Levantesi, Susanna, and Virginia Pizzorusso. "Application of machine learning to mortality modeling and forecasting." *Risks* 7.1 (2019): 26.
- 24) Levantesi, Susanna, Andrea Nigri, and Gabriella Piscopo. "Longevity risk management through Machine Learning: State of the Art." *Insurance Markets and Companies* 1.11 (2020): 11-20.
- 25) Loghi, Marzia, and Alessia D'Errico. "Il sistema informativo Health for All–Italia a supporto delle politiche sanitarie." *Ita sh* (2012): 36.
- 26) Natacha Brouhns, Michel Denuit, Jeroen K. Vermunt. "A Poisson log-bilinear regression approach to the construction of projected lifetables". *Insurance: Mathematics and Economics*. 2002;31(3):373–393
- 27) Nepomuceno, Marília R., et al. "The cross-sectional average inequality in lifespan (CAL+): A lifespan variation measure that reflects the mortality histories of cohorts." *Demography* 59.1 (2022): 187-206.
- 28) Nigri, Andrea, et al. "A deep learning integrated Lee–Carter model." *Risks* 7.1 (2019): 33.
- 29) Organizzazione Mondiale della Sanità. (2019). World Health Statistics 2019: Monitoring health for the SDGs.
- 30) Perla, Francesca, et al. "Time-series forecasting of mortality rates using deep learning." *Scandinavian Actuarial Journal* 2021.7 (2021): 572-598.
- 31) Posa Donato, Maggio Sabrina. "Elementi di demografia". (2017)
- 32) Schnürch, Simon, and Ralf Korn. "Point and interval forecasts of death rates using neural networks." *ASTIN Bulletin: The Journal of the IAA* 52.1 (2022): 333-360.
- 33) Treccani, Giovanni. "Enciclopedia Treccani." (*la mortalità*) (1998).
- 34) Vachon, Pierre J., and François Sestier. "Life expectancy determination." *Phys Med Rehabil Clin N Am* 24.3 (2013): 539-551.
- 35) Van Raalte, Alyson A., Isaac Sasson, and Pekka Martikainen. "The case for monitoring life-span inequality." *Science* 362.6418 (2018): 1002-1004.
- 36) Villegas, Andrés, Vladimir K. Kaishev, and Pietro Millossovich. "StMoMo: An R package for stochastic mortality modelling." *7th Australasian Actuarial Education and Research Symposium*. 2015.
- 37) Visco, Ignazio. "Lectio magistralis: Il rischio di longevità ei cambiamenti dell'economia." (2015).
- 38) World Health Organization. *World Health Statistics 2016 [OP]: Monitoring Health for the Sustainable Development Goals (SDGs)*. World Health Organization, 2016.