



**Studi e Analisi**

n. 1 - 2025

**ANALISI DELLE CAUSE DI  
RICALCOLO DELLE PRESTAZIONI  
COLLEGATE AL REDDITO:  
UN'INDAGINE SUI CONGUAGLI A  
DEBITO NELLE PENSIONI AI  
SUPERSTITI TRAMITE APPROCCIO  
"LASSO"**

Direzione centrale  
Studi e Ricerche



# ANALISI DELLE CAUSE DI RICALCOLO DELLE PRESTAZIONI COLLEGATE AL REDDITO: UN'INDAGINE SUI CONGUAGLI A DEBITO NELLE PENSIONI AI SUPERSTITI TRAMITE APPROCCIO "LASSO"<sup>1</sup>

## *Abstract*

Le prestazioni collegate al reddito sono erogate in via anticipata provvisoriamente, salva successiva verifica del diritto e misura sulla base dei dati reddituali consolidati a consuntivo messi a disposizione dalle Amministrazioni depositarie, fra cui in primis Agenzia delle Entrate. Ciò comporta che, a seguito delle verifiche reddituali effettuate a consuntivo sulla base dei dati consolidati provenienti da tutte le fonti dati a disposizioni (Agenzia delle Entrate, modelli RED resi dai cittadini, banche dati INPS, Casellario centrale delle pensioni), possano generarsi dei conguagli a debito, a credito o delle invarianze delle misure delle prestazioni collegate erogate. L'analisi dei dati di tali verifiche massive ha fatto emergere una maggiore propensione al ricalcolo delle pensioni ai superstiti. In particolare l'analisi dei dati presi in esame provenienti dalle campagne di verifica massiva reddituale RED 2020 (anno reddito 2019) e 2021 (anno reddito 2020), integrati con il Casellario centrale delle pensioni e le banche dati INPS, ha evidenziato che, con riferimento alle prestazioni collegate al reddito erogate nel 2020, i conguagli a debito delle pensioni ai superstiti erano circa il 4%, mentre nel 2021 sono saliti a quasi il 10% delle posizioni esaminate, rispettivamente l'1 e il 3% del totale delle prestazioni ai superstiti erogate. La mediana degli importi è però scesa da circa 400 euro l'anno a meno di 20 euro. Per quanto riguarda invece i conguagli a credito sono passati dal 4% delle posizioni esaminate nel 2020, a quasi il 20% nel 2021 e la media della distribuzione è intorno ai 60 euro per i conguagli rilevati nel 2020 e intorno ai 24 euro per quelli del 2021.

L'analisi effettuata supporta l'uso di modelli di probabilità e tecniche di regolarizzazione LASSO come strumento di *screening* per determinare il rischio di erogare pensioni al superstite di importo superiore o inferiore al dovuto. La metodologia proposta fornisce una stima del rischio che l'importo del suo trattamento non sia adeguato, a livello di singolo titolare di pensione ai superstiti. Tali stime possono essere utilizzate dall'Istituto per identificare i superstiti per cui il rischio di ricalcolo con conseguenti conguagli, a debito o a credito, è più elevato, e questo consente di prestare loro particolare attenzione già nella fase di definizione della pensione, pur nei limiti derivanti dalla mancanza di dati reddituali consolidati che potranno essere disponibili solo successivamente rispetto alla fase di liquidazione della prestazione.

---

<sup>1</sup> A cura di Giuseppe Pio Dachille, Monica Pia Cecilia Paiella (INPS, Direzione centrale Studi e Ricerche) e Cristiana Santulli (INPS, Direzione centrale Pensioni).

## 1. Introduzione

Le prestazioni collegate al reddito sono erogate dall'INPS in via anticipata provvisoriamente, salva successiva verifica del diritto e misura, ex art. 13, comma 2, della legge 30 dicembre 1991, n. 412, sulla base dei dati reddituali consolidati a consuntivo messi a disposizione dalle Amministrazioni depositarie, fra cui in primis Agenzia delle Entrate (ex art. 15, comma 1, del decreto-legge 1° luglio 2009, n. 78, convertito, con modificazioni, dalla legge 3 agosto 2009, n. 102). Ciò comporta che, a seguito delle verifiche reddituali effettuate a consuntivo sulla base dei dati consolidati provenienti da tutte le fonti dati a disposizione (Agenzia delle Entrate, modelli RED resi dai cittadini, banche dati INPS, Casellario centrale dei pensionati), possano generarsi dei conguagli a debito, a credito o delle invarianze delle misure delle prestazioni collegate erogate. L'analisi dei dati di tali verifiche massive ha fatto emergere una maggiore propensione al ricalcolo delle pensioni ai superstiti.

Il recupero dei pagamenti indebiti è uno dei principi fondamentali del Diritto delle obbligazioni: chiunque abbia versato una somma a qualcun altro senza una causa legittima ha il diritto di richiederne la restituzione. Questo principio, che a prima vista può sembrare semplice, può rivelarsi complesso da applicare nella pratica, indipendentemente dalla causa del pagamento indebito, sia essa dovuta a frode o a un semplice errore di una delle parti coinvolte.

Nel caso delle pensioni ai superstiti i pagamenti indebiti rappresentano una questione particolarmente delicata, perché coinvolgono persone anziane e vulnerabili, in situazioni di lutto. Inoltre, il recupero di tali pagamenti può risultare complicato a causa dell'età avanzata dei beneficiari. D'altro canto, anche un pagamento inferiore a quanto dovuto è da attenzionare proprio a causa della relativa fragilità dei beneficiari di queste prestazioni.

Questo studio si basa sui dati delle cosiddette campagne di verifica reddituale massiva delle prestazioni collegate al reddito che l'INPS conduce per conoscere i dati reddituali consolidati dei titolari di tali prestazioni a, tra cui i percettori di pensioni ai superstiti. Considerato che i redditi rilevanti possono variare nel corso degli anni, per garantire la correttezza degli importi erogati, l'INPS verifica la situazione reddituale rilevante ogni anno sulla base delle dichiarazioni rese dal pensionato: all'Agenzia delle Entrate tramite modelli 730/ REDDITI PF, attraverso il cosiddetto Modello RED all'INPS e le altre banche dati disponibili<sup>2</sup>. L'obiettivo dello studio è individuare i fattori che determinano la maggiore incidenza di conguagli con pagamenti di importo diverso da quanto dovuto in particolare per le pensioni ai superstiti.

---

<sup>2</sup> Il Modello RED viene compilato per rendere la dichiarazione della situazione reddituale rilevante, che il percettore di prestazioni correlate al reddito deve comunicare all'INPS annualmente. Non è obbligato a rendere tale dichiarazione chi ha già effettuato integralmente ad Agenzia delle Entrate la dichiarazione dei redditi rilevanti sulle prestazioni collegate in godimento.

Individuare tali fattori può aiutare a prevederne l'occorrenza e contribuire alla progettazione di strumenti per prevenirne almeno in parte il verificarsi, tenuto dei limiti fisiologici dovuti alla messa a disposizione dei dati rilevanti a consuntivo in una fase successiva alla liquidazione della prestazione. Inoltre, la metodologia può essere applicata anche ad altre prestazioni che presentano problematiche analoghe.

Il dataset utilizzato include circa un milione di pensioni ai superstiti erogate nel 2020 e nel 2021. Per ciascuna pensione ci sono informazioni sul trattamento stesso (ad es. importo, decorrenza e importo dell'eventuale conguaglio a debito/credito) e sulle caratteristiche del percettore della pensione e del dante-causa (es. reddito e occupazione, provincia di residenza). In base alle verifiche effettuate con le campagne RED 2020 e 2021, i conguagli a debito (con pagamenti superiori a quanto spettante) sono il 4% dei trattamenti erogati nel 2020 e quasi 10% di quelli erogati nel 2021, mentre i pagamenti inferiori (con conguagli a credito) sono il 4% e il 20% rispettivamente.

Per predire il rischio di ricalcolo di una prestazione con correlati conguagli, è possibile utilizzare un modello logistico, che stima la probabilità dell'evento come funzione lineare di un set di variabili che possono includere caratteristiche della prestazione, del beneficiario e, nel caso dei trattamenti al superstita, del dante causa, e loro interazioni. Per migliorare le capacità predittive e la robustezza del modello, data la potenziale numerosità delle variabili disponibili, utilizzeremo tecniche di *regolarizzazione*, provenienti dalla letteratura sul machine learning<sup>3</sup>. La principale caratteristica di queste tecniche è l'impiego di penalità per evitare l'*overfitting*, ovvero per limitare la complessità di un modello in presenza di molte variabili che potrebbe comportare l'eccessiva adattabilità del modello ai dati utilizzati per la stima, riducendone la robustezza e le capacità predittive su altri dataset. Inoltre, le tecniche di *regolarizzazione* risultano particolarmente utili quando alcune delle variabili a disposizione sono irrilevanti o collineari. La penalizzazione di queste variabili ne limita o azzerava il peso nel modello, riducendo la sensibilità delle stime a piccole variazioni nei dati e la probabilità di errore di classificazione su nuovi dataset.

Per prevedere il rischio di errori nella determinazione del trattamento al superstita e successivi conguagli a debito/credito, partiamo da un modello complesso con oltre 400 covariate, basate su dati presenti negli archivi INPS (Casellario centrale delle pensioni e Rapporti di lavoro dipendente del settore privato e pubblico). Attraverso l'impiego di tecniche Least Absolute Shrinkage and Selection Operator

---

<sup>3</sup> Un'alternativa alle tecniche di *regolarizzazione* sono le tecniche di *boosting*. Si tratta di metodi di apprendimento automatico che combinano più modelli deboli (cioè modelli che performano leggermente meglio del caso casuale, come piccoli alberi decisionali) per creare un modello complessivo forte, capace di ottenere elevate prestazioni predittive. Il termine "*boosting*" si riferisce all'idea di migliorare ("*boost*") la capacità predittiva complessiva iterativamente, per cui ogni modello successivo si concentra sulle osservazioni che i modelli precedenti hanno classificato in modo errato e i modelli deboli vengono combinati usando pesi che enfatizzano gli errori difficili da correggere. Esempi di tecniche di *boosting* sono AdaBoost, Gradient Boosting, e XGBoost.

(LASSO) identifichiamo il sottoinsieme di variabili che meglio predice il rischio di interesse e valutiamo la capacità del modello risultante di discriminare tra pensioni definite con assenza di conguagli post verifica reddituale e pensioni per cui sono necessari conguagli. Nel caso dei conguagli a debito, le covariate scendono a meno di 150, mentre nel caso dei conguagli a credito, il numero di variabili incluse varia tra le 150 e le 250 a seconda del dataset utilizzato (2020 o 2021). Per quanto riguarda le capacità previsive, il modello ottenuto risulta avere una capacità **più che accettabile** di discriminare tra pensioni al superstite definite con assenza di conguagli post verifica reddituale e prestazioni per cui si rilevano **conguagli a debito**, e una capacità previsiva accettabile nel caso di pensioni per cui si rilevano conguagli a credito. Inoltre, è emerso che i conguagli a debito e a credito sono più frequenti in presenza di contribuzione a specifiche tipologie di gestione pensionistica, diverse dal Fondo pensione dei lavoratori dipendenti.

Nel complesso, la metodologia sviluppata si mostra promettente come strumento di screening per determinare il rischio di erogare pensioni al superstite con conguagli a debito che a credito successivi alla verifica reddituale della misura effettivamente spettante. Essa, infatti, fornisce una stima del rischio per ciascun individuo (titolare di pensione al superstite) e può essere utilizzata dall'Istituto per identificare i superstiti per i quali il rischio di ricalcolo con conguagli successivi alla definizione del trattamento è più elevato, e questo consente di prestare loro particolare attenzione già nella fase di liquidazione della prestazione. Conoscere in anticipo il rischio individuale di futuri conguagli a debito/credito rappresenta un passo importante per ridurre il rischio di generazione di conguagli (a debito e a credito) riferiti alle pensioni al superstite erogate.

Prima di poter essere implementata e utilizzata a regime per prevedere ex ante i rischi sopra indicati, la metodologia va sottoposta a studi aggiuntivi per valutarne le capacità predittive e la precisione e a ulteriore validazione su altre campagne RED, per verificare l'accordo tra previsioni del modello e dati reali.

Il resto del documento è organizzato come segue. Nel secondo paragrafo, vengono descritte le principali caratteristiche delle pensioni al superstite, in termini di beneficiari, requisiti e misura delle prestazioni. Vengono anche descritte le Campagne RED di raccolta dei dati reddituali rilevanti che l'INPS conduce annualmente per verificare il diritto e la misura delle pensioni ai superstiti e, più in generale, le prestazioni collegate al reddito erogate in via anticipata provvisoriamente. Nel terzo paragrafo vengono presentati i dati. Nel quarto paragrafo, viene illustrata la metodologia econometrica. La lettura di questo paragrafo può essere omessa senza compromettere la comprensione della logica dell'analisi e dei risultati presentati nel quinto paragrafo. Nel sesto, si riassume l'analisi e si traggono le conclusioni.

## 2. Pensioni ai superstiti: caratteristiche, requisiti e validazione

La pensione ai superstiti è una prestazione previdenziale erogata ai familiari di un lavoratore o di pensionato deceduto<sup>4</sup>. Ha lo scopo di garantire un sostegno economico ai congiunti del defunto, riconoscendo loro una percentuale della pensione che il deceduto percepiva o avrebbe avuto diritto a percepire. Spetta ai familiari del titolare di una qualsiasi prestazione previdenziale (es. pensione di vecchiaia, anticipata, di inabilità, di anzianità) ad eccezione del solo assegno ordinario di invalidità in quanto questo, secondo quanto disposto dalla Legge n. 222/1984, non è reversibile. I superstiti hanno anche diritto alla pensione supplementare di reversibilità qualora il defunto sia titolare anche di pensione supplementare diretta.

I familiari che possono beneficiare della pensione ai superstiti comprendono il coniuge, i figli minori, o studenti fino a 26 anni, o, senza limiti di età, se inabili al lavoro e a carico del defunto, i genitori con più di 65 anni e a carico del defunto, fratelli e sorelle se inabili al lavoro e a carico del defunto e i nipoti minori o maggiorenni inabili al lavoro e a carico del defunto.

La pensione ai superstiti viene calcolata come percentuale della pensione percepita dal defunto o che gli spetterebbe nel caso di decesso di un lavoratore, come segue:

<b>Superstite</b>	<b>%</b>
Coniuge senza figli	60%
Coniuge con un figlio <sup>5</sup>	80%
Coniuge con due o più figli	100%
Un figlio	70%
Due figli	80%
Tre o più figli	100%
Ciascun genitore o fratello o sorella <sup>6</sup>	15%

<sup>4</sup> Nel caso di decesso prima del pensionamento, il trattamento al superstite viene concesso se il lavoratore, al momento del decesso, aveva 15 anni di contribuzione oppure 5 anni di contribuzione, di cui 3 nell'ultimo quinquennio.

<sup>5</sup> Ai figli sono equiparati i nipoti minori o maggiorenni inabili, se a carico dell'ascendente al momento della sua morte.

<sup>6</sup> I genitori e i fratelli e sorelle possono percepire la pensione al superstite solo in mancanza del coniuge e/o dei figli del deceduto.

La pensione ai superstiti viene ridotta se il beneficiario possiede altri redditi, secondo le seguenti modalità:

<b>Altri redditi</b>	<b>% riduzione</b>
Fino a 3 volte la pensione minima	Nessuna
Tra 3 e 4 volte la pensione minima	25%
Tra 4 e 5 volte la pensione minima	40%
Oltre 5 volte la pensione minima	50%

Nota: nel 2024 la pensione minima è fissata a 598,61 euro.

La riduzione non si applica quando tra i beneficiari ci sono figli minori o maggiorenni studenti o inabili e per le pensioni di reversibilità con decorrenza antecedente al 1° settembre 1995 (che conservano il trattamento più favorevole in pagamento al 31/08/1995, con riassorbimento sui futuri miglioramenti). In tutti i casi ove l'importo della pensione al superstite è subordinato all'entità dei redditi del beneficiario, l'INPS è tenuto a verificarlo, così come avviene per tutte le prestazioni subordinate al reddito, in base alla Legge n. 412 del 1991 (art. 13, comma 2) in combinato disposto con l'art. 35, comma 10 bis, del D.L. n. 207 del 2008. Allo scopo, il percettore è tenuto a dichiarare all'Istituto i propri dati reddituali rilevanti ogni anno, attraverso la presentazione del modello 730 o REDDITI PF ad Agenzia delle Entrate, ovvero mediante dichiarazione diretta all'Istituto con il modello RED. L'INPS procede quindi all'elaborazione massiva, centralizzata, dei dati raccolti relativi alla situazione reddituale rilevante per effettuare la verifica finalizzata alla corretta determinazione del diritto e della misura della prestazione che l'Istituto ha erogato provvisoriamente in via anticipata. In caso di mancata comunicazione dei redditi, l'INPS provvede a sollecitare i soggetti inadempienti ed in mancanza di risposta, procede alla sospensione delle prestazioni collegate al reddito ed alla successiva rideterminazione in via definitiva per l'anno reddito richiesto, qualora non pervenga oltre il limite fissato dall'Istituto.

Per la verifica reddituale massiva delle prestazioni collegate al reddito, oltre alla dichiarazione del percettore, l'Istituto si avvale del sistema informativo basato sull'interscambio di dati fra le pubbliche amministrazioni e del sistema informativo interno basato sui flussi provenienti dalle altre Direzioni Centrali. Dal confronto analitico dei redditi rilevanti, il diritto e la misura di una prestazione possono essere confermati o rideterminati, con conseguenti conguagli a credito o a debito. In caso di incongruenze o incompletezza dei dati reddituali rilevanti, la prestazione viene rideterminata, e in alcuni casi particolari alimenta le liste delle segnalazioni di posizioni da lavorare a cura delle sedi.

### 3. Il Casellario delle pensioni e le campagne RED 2020 e 2021

Le cosiddette Campagne RED vengono condotte annualmente dall'Istituto e prendono il nome dal modello di autodichiarazione dei redditi che sono tenuti a compilare i titolari di prestazioni collegate al reddito che non presentano il modello 730 o REDDITI PF, e coloro che non comunicano integralmente ad Agenzia delle Entrate la situazione reddituale incidente sulle prestazioni collegate in godimento, perché ad esempio hanno anche altri redditi influenti sulle prestazioni non inclusi tra quelli che devono essere comunicati all'Amministrazione finanziaria in sede di dichiarazione dei redditi. La platea di riferimento per le Campagne RED esaminate sono i titolari delle prestazioni collegate al reddito della gestione privata<sup>7</sup>. Attraverso la Campagna di verifica reddituale RED vengono prese in considerazione le informazioni reddituali rilevanti diverse da pensioni riferite all'anno precedente a quello di erogazione, mentre per i redditi da pensione vengono valorizzati i dati presenti nel Casellario Centrale pensioni con riferimento all'anno di erogazione.

Ai nostri fini, siamo partiti dai dati delle campagne RED 2020 e 2021, ordinarie e relativi solleciti, sui redditi 2019 e 2020 dei titolari di pensioni ai superstiti, che si sono chiuse rispettivamente il 21 marzo 2022 e il 28 febbraio 2023. Abbiamo quindi incrociato i risultati delle verifiche con i dati relativi alle pensioni ai superstiti (da assicurato e da pensionato) presenti nel Casellario centrale delle pensioni, escludendo la Gestione dei dipendenti pubblici, i trattamenti con decorrenza anteriore al 1° settembre 1995 ed i titolari di pensioni ai superstiti che sono deceduti prima della conclusione della campagna RED (per es., per la RED 2020, abbiamo escluso i deceduti nel 2020 e 2021). L'esito dell'incrocio, a livello di prestazione, è riportato nella Tabella 1. I titolari di circa un terzo delle prestazioni hanno partecipato alla campagna RED. I titolari degli altri due terzi non vi hanno partecipato e, tra questi, vi sono i titolari di prestazioni non soggetti all'obbligo di dichiarazione della situazione reddituale (es. titolari di pensione al superstite con minori a carico). Le campagne di verifica possono dare luogo a conguagli a debito o a credito del titolare della prestazione. L'incidenza dei conguagli sul totale delle posizioni esaminate varia notevolmente da un anno all'altro: nella campagna del 2020, ammontano ciascuno a circa il 4% delle posizioni esaminate; nella campagna del 2021, invece, ammontano rispettivamente a quasi il 10% e a circa il 20% delle posizioni esaminate. Se rapportati al totale delle prestazioni al superstite considerate, ovvero includendo anche le prestazioni non esaminate, le percentuali dei conguagli a debito e a credito rilevati nel 2020 scendono al 1% ciascuno, mentre quelli del 2021 scendono al 3 e 6% rispettivamente.

---

<sup>7</sup> Fino al 2021, le verifiche delle prestazioni pensionistiche collegate al reddito della gestione pubblica sono state effettuate dai sistemi proprietari che le gestiscono. Dall'anno 2022, l'INPS liquida anche le prestazioni della gestione pubblica con i sistemi integrati (nuova IVS), per cui dall'anno reddito 2022 anche le nuove liquidate con i sistemi integrati sono sottoposte a verifica attraverso le Campagne RED (cioè con le medesime modalità previste per le pensioni di gestione privata). I trattamenti già vigenti gestiti dai sistemi proprietari della Gestione dipendenti pubblici continuano a seguire un autonomo iter di verifica, nelle more del processo di migrazione.

Tabella 1 – Prestazioni ai superstiti

	2020	2021
Posizioni invariate	889.800	736.727
Posizioni con conguaglio a debito	38.904	100.512
Posizioni con conguaglio a credito	43.141	203.034
Posizioni esaminate (rilevate nel Casell. e in C. RED)	971.845	1.040.273
Posizioni Casell., non rilevate in C. RED*	2.123.115	1.867.430
<i>di cui successivamente eliminate</i>	<i>132.407<sup>8</sup></i>	<i>4.823<sup>9</sup></i>

Nota: la tabella riporta nella prima (seconda) colonna i risultati dell'incrocio delle posizioni esaminate attraverso la campagna RED 2020 con le prestazioni al superstite del Casellario C.P. del 2020 (2021). (\*) Si tratta delle prestazioni al superstite non esaminate attraverso la campagna RED 2020 (2021), per le quali presumibilmente non è stato ricostituito il reddito, e nell'ultima riga quelle che, tra queste, sono state eliminate negli anni successivi.

Tabella 2 – Titolari di prestazioni ai superstiti

	2020	2021
Titolari con posizioni invariate	838.947	682.612
Titolari con posizioni con conguaglio a debito	38.871	100.319
Titolari con posizioni con conguaglio a credito	43.112	202.886
Titolari esaminati (rilevati in Casell. e in C. RED)	920.704	985.817
Posizioni Casellario, non rilevate in C. RED	2.044.336	1.793.970
<i>di cui successivamente eliminate</i>	<i>127.291<sup>10</sup></i>	<i>4.724<sup>11</sup></i>

Nota: vedi nota della Tabella 1.

La Tabella 2 riporta l'esito dell'incrocio per titolare di pensione al superstite. Nel 2020 e 2021, circa il 9% dei titolari di questa tipologia di trattamento percepisce più di una pensione al superstite per cui le campagne di controllo possono dare esiti diversi per le varie pensioni. Per questo motivo, la somma dei percettori con posizioni a credito, debito e invariate è superiore al totale dei partecipanti alle campagne RED. Ai fini di questa tabella, aggregiamo le pensioni al superstite per titolare e aggregiamo anche eventuali conguagli a debito tra loro e quelli a credito tra loro, per cui i titolari di più di una prestazione

<sup>8</sup> Pensioni eliminate nel 2022 e 2023.

<sup>9</sup> Pensioni eliminate nel 2023, anno in cui si è conclusa la campagna RED. Altre potrebbero essere state eliminate nel 2024 per cui i dati non sono ancora definitivi.

<sup>10</sup> Pensioni eliminate nel 2022 e 2023.

<sup>11</sup> Pensioni eliminate nel 2023, anno in cui si è conclusa la campagna RED. Altre potrebbero essere state eliminate nel 2024 per cui non abbiamo ancora i dati.

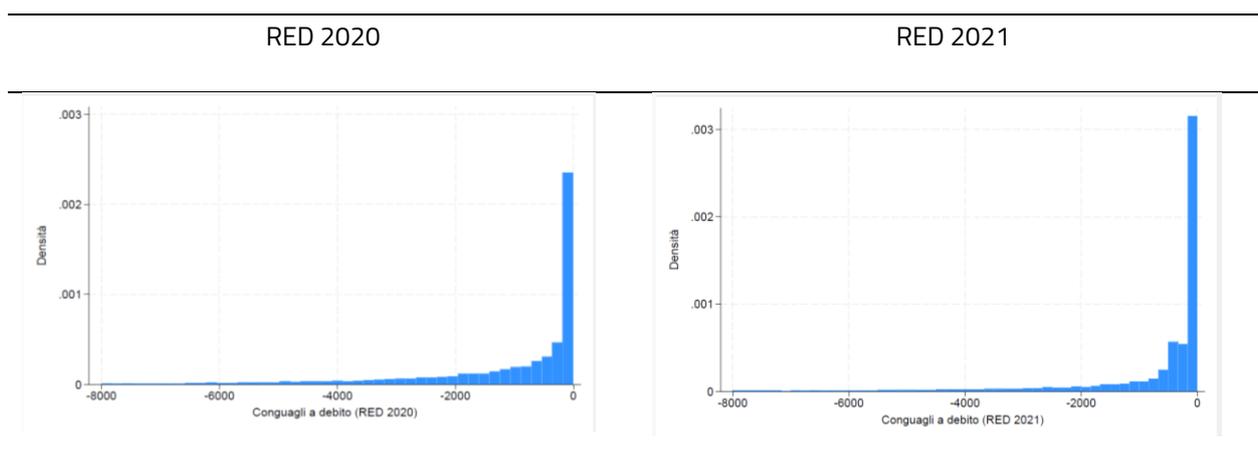
possono avere una posizione con conguaglio a debito e una a credito e/o una invariata, ma non due o più prestazioni con conguaglio dello stesso segno.

Il resto dell'analisi è a livello di percettore di prestazione e la Tabella 3 riporta alcune caratteristiche dei titolari di pensioni al superstite che hanno partecipato alle campagne RED 2020 e 2021 distinguendo in base all'esito della campagna. Oltre a differenze notevoli nell'incidenza dei conguagli tra i due anni, ci sono anche differenze in termini di caratteristiche dei titolari di pensione.

In generale, i conguagli a debito sono più frequenti tra le pensioni con percettori di sesso maschile, residenti nel Nord del paese, per pensioni con decorrenza di 1 anno o meno, per pensioni erogate dalle gestioni Artigiani e Commercianti e da altre gestioni (diverse dal Fondo pensione dei lavoratori dipendenti) e tra coloro che percepiscono anche altri trattamenti. L'importo medio delle pensioni con conguaglio a debito è di 5-700 euro lordi superiore alla media delle pensioni al superstite.

La distribuzione dei conguagli a debito è molto eterogenea, concentrata su valori esigui, soprattutto quelli della Campagna RED 2021, ma con code molto lunghe (Grafico 1). La distribuzione dei conguagli della RED 2020 ha mediana pari a 400 euro e deviazione standard di circa 3.600. Il 25esimo e 75esimo percentile sono rispettivamente pari a 50 e 1900 euro. Conguagli a debito di minore entità si rilevano nella campagna 2021. La distribuzione ha mediana pari a 19 euro e deviazione standard di circa 3.300. Il 25esimo e 75esimo percentile sono rispettivamente pari a 1 e 400 euro.

Grafico 1 – I conguagli a debito

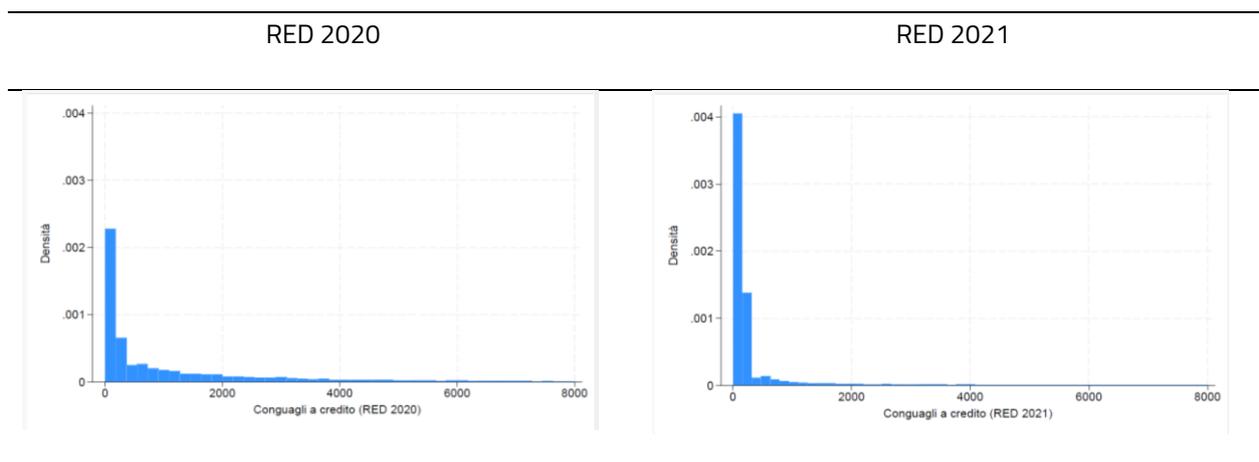


Nota: La distribuzione è basata sugli esiti delle Campagne RED.

Per quanto riguarda invece i conguagli a credito, l'incidenza è massima tra i titolari di pensioni al superstite che percepiscono altre pensioni e nel caso di pensioni al superstite liquidate dalle gestioni Artigiani e Commercianti e da altre gestioni (diverse dal Fondo pensione lavoratori dipendenti). Inoltre,

ma solo nel 2020, rileva la decorrenza per cui l'incidenza dei conguagli a credito è massima tra le pensioni liquidate 2 e 3 anni prima della Campagna RED, mentre, nel 2021, conta l'area di residenza del percettore, per cui l'incidenza è massima tra i residenti nel Nord. In termini di importo, il valore mediano dei conguagli a credito per il 2020 e il 2021 è rispettivamente pari a 62 e 24 euro annui, con delle distribuzioni molto "disperse" (deviazione standard di 2.700 e 1.900) e il 75esimo percentile di 900 e 550 euro (Grafico 2).

Grafico 2 – I conguagli a credito



Nota: La distribuzione è basata sugli esiti delle Campagne RED.

Tabella 3 – Statistiche descrittive per titolare di prestazione al superstite (RED 2020 e 2021)

	Partecipanti <u>2020</u>				Partecipanti <u>2021</u>			
	Totale	conguaglio a debito	conguaglio a credito	posizione invariata	Totale	conguaglio a debito	conguaglio a credito	posizione invariata
Età	78,63	<b>75,63</b>	76,66	78,86	78,4	77,02	79,72	78,21
Genere (donne)	0,86	<b>0,80</b>	0,85	0,87	0,87	<b>0,79</b>	0,89	0,88
Decorrenza: ≤1 anno	0,11	<b>0,29</b>	0,07	0,10	0,11	<b>0,17</b>	0,09	0,11
Decorrenza: 2-3 anni	0,27	0,28	<b>0,37</b>	0,26	0,24	0,25	0,23	0,24
Decorrenza: 4-5 anni	0,25	0,17	0,23	0,25	0,22	0,28	0,24	0,22
Decorrenza: >5	<u>0,37</u>	<b><u>0,26</u></b>	<u>0,33</u>	<u>0,38</u>	<u>0,43</u>	<b><u>0,38</u></b>	<u>0,44</u>	<u>0,43</u>
<i>Totale</i>	<i>1,00</i>	<i>1,00</i>	<i>1,00</i>	<i>1,00</i>	<i>1,00</i>	<i>1,00</i>	<i>1,00</i>	<i>1,00</i>
Importo lordo annuo	€9 846	<b>€10 558</b>	€9 800	€9 818	€9 534	<b>€10 011</b>	€9 216	€9 577
<i>(deviazione standard)</i>	<i>(6 021)</i>	<i>(7 987)</i>	<i>(6 773)</i>	<i>(5 879)</i>	<i>(6 081)</i>	<i>(8 416)</i>	<i>(5 635)</i>	<i>(5 821)</i>
Gestione: FPLD	0,58	<b>0,48</b>	<b>0,50</b>	0,58	0,57	<b>0,52</b>	0,55	0,58
Gestione: Art, e Comm,	0,24	<b>0,32</b>	0,28	0,23	0,25	<b>0,27</b>	0,26	0,24
Altre gestioni	0,18	0,20	0,22	0,19	0,18	0,21	0,19	0,18
<i>Totale</i>	<i>1,00</i>	<i>1,00</i>	<i>1,00</i>	<i>1,00</i>	<i>1,00</i>	<i>1,00</i>	<i>1,00</i>	<i>1,00</i>
Titolare di altre pensioni	0,50	<b>0,59</b>	0,53	0,50	0,50	<b>0,62</b>	<b>0,55</b>	0,46
Residente nel Nord	0,51	0,53	0,49	0,51	0,52	<b>0,56</b>	0,51	0,51
Residente nel Centro	0,20	0,21	0,20	0,20	0,20	0,21	0,20	0,2
Residente nel Mezzogiorno	0,29	0,26	0,31	0,29	0,29	0,23	0,29	0,29
# osservazioni	920 704	38 871	43 112	838 947	985 817	100 319	202 886	682 612

Tabella 4 – Statistiche descrittive per dante causa delle prestazioni al superstite (RED 2020 e 2021)

	Dante causa campagna <u>2020</u>				Dante causa campagna <u>2021</u>			
	Totale	conguaglio a debito	conguaglio a credito	posizione invariata	Totale	conguaglio a debito	conguaglio a credito	posizione invariata
Età alla morte	78,06	<b>75,63</b>	76,13	78,28	77,73	75,95	78,46	77,77
Genere (donne)	0,14	<b>0,20</b>	0,15	0,13	0,13	<b>0,21</b>	0,11	0,12
Anzianità contrib. (settimane)	1.448,23	<b>1.467,74</b>	1.447,62	1.447,49	1.475,04	<b>1.442,59</b>	<b>1.438,78</b>	1.491,59
Importo lordo pensione (€2022)	6.026,24	<b>6.814,73</b>	<b>6.414,18</b>	5.971,90	6.201,70	<b>7.389,14</b>	6.019,67	6.100,17
Ultimo reddito da lavoro (€2022)	31.741,61	<b>41.495,91</b>	35.331,74	29.332,32	30.984,40	<b>48.194,38</b>	<b>37.177,22</b>	24.520,02
# osservazioni	920.615	38.870	43.108	838.863	983.685	100.293	202.842	682.471

Infine, la Tabella 4 riporta alcune caratteristiche relative ai danti causa, ovvero gli assicurati originari. Conguagli a debito sulla pensione al superstite sono più frequenti quando il dante causa è una donna e ha una pensione di importo superiore alla media (differenza di circa 1000 euro rispetto alle pensioni dei danti causa a cui si associano pensioni al superstite "invariate"). Dall'incrocio delle informazioni sugli assicurati con quelle degli archivi dei rapporti di lavoro privato, disponibili dal 1974, e pubblico, disponibili dal 2014, sono state ricostruite le loro carriere lavorative e il loro reddito. Come per le loro pensioni, anche i redditi da (ultimo) lavoro dei danti causa di pensioni al superstite con conguagli a debito sono significativamente superiori alla media: la differenza rispetto ai redditi dei danti causa di "pensioni invariate" è di circa 12.000 euro nella campagna del 2020 e di 24.000 euro in quella del 2021. Per quanto riguarda le pensioni con conguagli a credito, la differenza più significativa in termini di caratteristiche dei danti causa nel 2020 riguarda l'importo della loro pensione che è di circa 500 euro superiore a quella degli assicurati associati a "pensioni invariate", e nel 2021 il loro (ultimo) reddito che è di circa 12.000 euro superiore a quella degli assicurati associati a "pensioni invariate".

#### 4. La metodologia empirica

Come già accennato, l'obiettivo dello studio è individuare una metodologia che individui le prestazioni per cui è maggiore il rischio che le somme erogate a titolo di pensione al superstite differiscano da quanto dovuto. Il modello più comune per problemi binari (es. pagamento indebito: si/no) è la regressione logistica che fornisce probabilità predittive chiare e può essere rappresentato nel seguente modo:

$$\text{logit}(P_{i,t}) = \ln\left(\frac{P_{i,t}}{1 - P_{i,t}}\right) = \beta' X_{i,t},$$

dove  $P_{i,t}$  è la probabilità che il titolare di una pensione al superstite riceva nell'anno t un pagamento diverso da quanto dovuto, ovvero  $P_{i,t} = \text{Prob}(y_i = 1)$ .  $y_i$  è una variabile dicotomica che assume valore 1 in presenza di un conguaglio a debito; altrimenti ha valore 0. La variabile dipendente  $\text{logit}(P_{i,t})$  è una trasformazione logistica di  $P_{i,t}$ .  $\beta = (\beta_0, \beta_1, \beta_2, \dots, \beta_p)$  è il vettore dei coefficienti delle variabili del modello.  $X_{i,t}$  è il vettore di fattori che predicano l'errore e include, in teoria, caratteristiche della pensione (es. importo, decorrenza...), del titolare del trattamento (es. genere, titolarità di altri trattamenti pensionistici...) e del dante causa (es. gestione, anzianità contributiva...).

Nel caso di modelli complessi, con molte variabili indipendenti, quando la regressione logistica ha problemi di convergenza, un'alternativa è il ricorso alla regressione lineare con variabile dipendente che assume valore 1/0. Sia nel caso di modello logistico che di quello lineare, in presenza di molte covariate,

risultano particolarmente utili le tecniche di regolarizzazione, che forniscono una selezione automatica delle variabili, e riducono il rischio di *overfitting*, contribuendo alla generalizzazione del modello su altri dati. La tecnica di regolarizzazione che utilizziamo è il Least Absolute Shrinkage and Selection Operator (LASSO) che aggiunge al modello un set di penalità proporzionale al valore assoluto dei coefficienti<sup>12</sup>. Le penalità del LASSO tendono a ridurre il valore dei coefficienti e possono anche azzerarne alcuni. Ciò è particolarmente utile in contesti ad alta dimensionalità, perché si realizza una selezione delle variabili per cui vengono incluse solo quelle che hanno un contributo significativo. Anche in presenza di multicollinearità, il LASSO seleziona la variabile più rappresentativa e penalizza le altre. Inoltre, azzerando i coefficienti di alcune variabili irrilevanti, viene meno l'instabilità causata da queste ultime. Tutto questo contribuisce alla stabilità del modello, garantendo stime più affidabili. Oltre a questo, modelli stabili tendono a non sovradattarsi ai dati di addestramento (*overfitting*) e a funzionare meglio su nuovi dati. Infine, il sistema di penalità limita la tendenza dei coefficienti ad assumere valori molto grandi e assicura così che le variabili contribuiscano in modo più uniforme al modello, riducendo la sensibilità a perturbazioni nei dati o a correlazioni tra le variabili.

Lo stimatore LASSO<sup>13</sup> del vettore dei coefficienti  $\beta = (\beta_0, \beta_1, \beta_2, \dots, \beta_p)$  si ottiene dall'ottimizzazione della seguente funzione di verosimiglianza:

$$L(\beta_1, \beta_2, \dots, \beta_p) = \sum_{i=1}^n \ln [P(x_i; \beta)^{y_i} + (1 - P(x_i; \beta))^{1-y_i}] + \lambda \sum_{j=1}^p |\beta_j|, \quad \text{con: } \lambda \geq 0.^{14}$$

Il termine  $\lambda \sum_{j=1}^p |\beta_j|$  rappresenta la penalità<sup>15</sup> per effetto della quale i coefficienti delle variabili indipendenti ridondanti, che non apportano cioè informazione incrementale utile al modello, generalmente assumono valori vicini o pari a zero. Ciò equivale a ridurre la complessità del modello, dove la complessità è determinata dal numero di variabili indipendenti e relativi coefficienti che definiscono il modello, e, in questo modo, la regressione LASSO può aiutare a identificare le variabili più rilevanti rendendo il modello più semplice e interpretabile. Il parametro  $\lambda$  controlla l'ammontare della penalità. Se  $\lambda=0$  lo stimatore coinciderà con lo stimatore di massimo-verosimiglianza, mentre se  $\lambda \rightarrow \infty$

<sup>12</sup> In base al tipo di penalità, si distingue tra modelli:

- LASSO: Promuove la 'sparsità', eliminando completamente alcune caratteristiche;
- Ridge: Riduce i parametri molto grandi senza annullarli, rendendo il modello più stabile;
- Elastic net: Combina LASSO e Ridge per ottenere un bilanciamento tra 'sparsità' e stabilità.

<sup>13</sup> Tibshirani, R., "Regression Shrinkage and Selection via the Lasso", Journal of the Royal Statistical Society. Series B (Methodological), Vol. 58, No. 1, 1996, pp. 267-288.

<sup>14</sup> Nel caso di modello di probabilità lineare, la funzione obiettivo è:

$$L(\beta_1, \beta_2, \dots, \beta_p) = \sum_{i=1}^n (y_i - \sum_{j=1}^p \beta_j x_{i,j})^2 + \lambda \sum_{j=1}^p |\beta_j|, \quad \text{con: } \lambda \geq 0.$$

<sup>15</sup> La presenza di questo termine è fonte di bias, il cui impatto negativo sulle stime può tuttavia essere inferiore al guadagno in termini di minore varianza, migliorando quindi l'accuratezza complessiva del modello. Queste tecniche, infatti, ottimizzano il trade-off tra bias e varianza, al fine di ottenere un modello robusto che conduca a predizioni accurate su dataset diversi da quello utilizzato per la stima.

l'impatto della penalità aumenta ed i coefficienti tenderanno a zero. La scelta del parametro di regolarizzazione  $\lambda$  è un elemento cruciale nella stima del modello., che determina le variabili da includere nel modello e quelle con coefficiente zero. Il valore di  $\lambda$  viene generalmente determinato attraverso un processo di *cross-validation*, che consiste nella stima iterativa del modello, che assegna ad ogni iterazione valori diversi a  $\lambda$  e lo stima utilizzando combinazioni diverse di dati. Nello specifico utilizziamo una procedura di validazione incrociata su 10 campioni (*10-fold cross-validation*) che prevede la suddivisione del dataset in 10 parti. Il modello viene stimato sul 90% dei dati e testato sul restante 10% e il processo viene ripetuto 10 volte, utilizzando ciascun *fold* per la validazione una volta. Ciò garantisce che ogni dato venga utilizzato sia per la stima che per la validazione. La validazione incrociata a 10 *fold* media le prestazioni su più *fold*, riducendo l'impatto di una specifica suddivisione e portando a una stima più stabile.

Per valutare la bontà del modello, ovvero la sua capacità di predire un conguaglio a debito/credito, utilizziamo la cosiddetta *area under the curve* (AUC) che misura la capacità del modello di distinguere tra superstiti i cui trattamenti vengono liquidati correttamente e trattamenti per cui la successiva campagna RED rileva errori e quindi la necessità di operare conguagli a debito/credito. Il valore dell'AUC misura questa capacità e varia tra un valore soglia<sup>16</sup> del 50%, che indica incapacità di discriminare tra pensioni al superstite corretti e non-corrette, valori dal 70% all'80% – accettabile, dall'80% al 90% – eccellente, e 90% – prestazioni eccezionali. Anche la stima dell'AUC del modello finale si basa sulla validazione incrociata su 10 campioni casuali<sup>17</sup>.

## 5. I risultati dell'analisi

Come spiegato, l'obiettivo dell'analisi è utilizzare i dati sull'incidenza dei conguagli a debito e a credito delle pensioni al superstite rilevati nelle Campagne RED 2020 e 2021 per individuare una metodologia che consenta di prevederne ex ante il rischio di generazione degli stessi sulla base dei dati già a disposizione dell'Istituto. Allo scopo, utilizziamo un modello di probabilità e il LASSO per selezionare le covariate che hanno maggiore potere predittivo. Per la predizione di conguagli non-nulli partiamo infatti da una batteria molto ampia di variabili e loro interazioni. Il totale delle covariate (variabili e loro interazioni) disponibili è superiore a 400 e include: polinomi di secondo grado negli importi della pensione al superstite, di eventuali altre pensioni del superstite e della pensione del dante causa, a

---

<sup>16</sup> Cfr.: Steyerberg EW et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*. 2010;21(1):128–38; and: Hosmer DW Jr et al. *Applied logistic regression*. Hoboken (NJ): John Wiley & Sons: 2013.

<sup>17</sup> Cfr.: Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J*. 2014;35 (29):1925–31; e Brathwaite R et al. Predicting the individualized risk of poor adherence to ART medication among adolescents living with HIV in Uganda: the Suubi+Adherence study. *Journal of the International AIDS Society* (24) 2021.

prezzi costanti (in euro 2021); l'anzianità contributiva, l'anno di decorrenza e una batteria di dummies per le gestioni della pensione al superstite e di altre eventuali pensioni del superstite; l'eventuale ultimo reddito da lavoro del superstite, a prezzi costanti, distinguendo tra reddito da lavoro dipendente nel settore privato e pubblico; il genere e la sua età al momento della liquidazione della pensione; l'anno di decorrenza della pensione del dante causa, e il suo ultimo reddito da lavoro, a prezzi costanti, distinguendo tra reddito da lavoro dipendente nel settore privato e pubblico e la sua età al momento del decesso; e una batteria di dummies per la provincia di residenza del superstite. Il LASSO (Least Absolute Shrinkage and Selection Operators) ne riduce (Shrinkage) il numero, selezionando (Selection) quelle il cui potere esplicativo è più elevato. In questo modo, riduce la complessità del modello che viene ad includere solo le covariate più rappresentative e che più contribuiscono a spiegare l'evento di interesse.

Le Tabelle 5 e 6 mostrano l'esito del LASSO e riportano i coefficienti standardizzati delle prime 20 variabili selezionate, ordinate in termini di importanza predittiva nel modello di previsione dei conguagli a debito (Tabella 5) e a credito (Tabella 6)<sup>18</sup>. La prima colonna riporta i risultati delle stime basate sulla campagna RED 2020, mentre la seconda colonna riporta quelli della campagna RED 2021. I coefficienti di regressione mostrati sono "penalizzati" in quanto sono stati calcolati applicando la penalizzazione LASSO per ridurre l'*overfitting*. Sono quindi *biased* e non riflettono le vere associazioni a livello di popolazione. Pertanto, l'attenzione non dovrebbe essere rivolta all'interpretazione dei singoli coefficienti delle variabili, ma piuttosto a come il modello si comporta nel complesso considerando la combinazione di tutte le covariate. Nei modelli di previsione dei conguagli a debito (Tabella 5), sui dati del 2020 e 2021, l'algoritmo ha selezionato rispettivamente 149 e 112 covariate (su oltre 400), mentre in quello di previsione dei conguagli a credito (Tabella 6) 149 e 244. Per quanto l'interpretazione dei coefficienti esuli dalle finalità predittive della nostra analisi, è interessante osservare come le caratteristiche che sono associate a coefficienti standardizzati di *magnitudo* più elevata in valore assoluto siano relative alla presenza di gestioni associate a rapporti di lavoro non dipendente, come la cassa ENPAM e la Gestione Separata. Inoltre, anche la titolarità di altre pensioni, oltre a quelle da superstite, sembra essere un fattore rilevante per il verificarsi di un indebito. Un'altra caratteristica che sembra avere alto potere predittivo, poi, è la breve decorrenza della pensione da superstite, come si evince dalla presenza tra le covariate associate a coefficienti più "grandi" della dummy "recente" e delle sue interazioni. Per la presenza di conguagli a credito, infine, sembrano rilevare la contribuzione a fondi integrativi e sostitutivi dell'AGO.

---

<sup>18</sup> Le tecniche LASSO standardizzano automaticamente i coefficienti e questo consente di confrontare direttamente l'importanza relativa delle variabili predittive, indipendentemente dalla loro scala originale.

Tabella 5 – La previsione delle pensioni con conguaglio a debito

	2020		2021
Selected lambda	.0005519		.0018067
# of covariates	408		405
# of non-zero coefficients	149		112
(Recente* = 1) x ENPAM	0.0156	Gestione (altra pens.): ENPAM	0.0650
(Recente = 1) x Importo altra pens.	0.0143	Gestione: Parasubordinati	0.0231
Gestione: Ex-dipendenti Inps	0.0127	Gestione (altra pens.): Parasubordinati	0.0169
(Femm. = 1) x Importo altra pens.	0.0098	Gestione (altra pens.): Ex-dipendenti Inps	0.0168
(Femm. = 1) x Redd. lav. dipend. (pubbl.)	0.0084	(Recente = 1) x Importo altra pens.	0.0161
Anni dalla liquidazione al quadrato	0.0076	Anno decorrenza altra pens.	0.0156
(Femm. = 0) x Anzianità contrib.	0.0073	Gestione (altra pens.): FPLD	-0.0135
Età del superst. alla liquidazione	-0.0069	Gestione (altra pens.): Elettrici	0.0118
(Recente = 1) x Anz. contrib. (altra pens.)	0.0065	(Femm. = 1) x Anzianità contrib.	-0.0117
(Recente = 1) x FPLD	-0.0065	Gestione (altra pens.): Ipost	0.0104
Redd. lav. dipend. (priv.)	0.0064	Gestione (altra pens.): Totalizzate	0.0103
(Femm. = 1) x FPLD (altra pens.)	-0.0064	Importo pens. dante causa	0.0102
(Recente = 1) x Importo pens. superst.	0.0062	(Femm. = 1) x Ipost (altra pens.):	0.0102
Anni dalla liquidazione	-0.0055	(Femm. = 1) x Importo pens. superst.	-0.0097
Gestione: Ex-Inpdai	0.0055	Gestione (altra pens.): CDCM post 90	-0.0092
(Femm. = 1) x Redd. lav. dipend. (priv.)	0.0053	(Femm. = 1) x Ex-Inpdai	0.0091
(Femm. = 0) x Importo altra pens.	-0.0045	Gestione: ENPAM	0.0087
(Femm. = 1) x CDCM post 90 (altra pens.)	-0.0039	Gestione (altra pens.): Ferrovieri	0.0081
Gestione altra pens.: Commercianti	0.0037	(Femm. = 1) x Redd. da lav. dipend. (pubbl.)	0.0077
(Femm. = 1) x Totalizzazione (altra pens.)	0.0035	(Femm. = 0) x ENPAM (altra pens.):	0.0077
Redd. lav. dipend. (pubbl.)	0.0034	Importo pens. superst.	-0.0073
(Femm. = 1) x Anzianità contrib.	-0.0034	Reddito da lav. dipendente (pubblico)	0.0072
(Recente = 1) x Telefonici (altra pens.)	0.0033	Importo altra pens.	0.0071
Gestione: Parasubordinati	0.0032	Gestione (altra pens.): Telefonici	0.0071
(Recente = 1) x Ex-Inpdai	0.0031	Provincia MI	0.0063
(Recente = 1) x CDCM post 90 (altra pens.)	-0.0030	(Recent = 0) x Ex-Inpdai (altra pens.)	0.0059
(Recente = 1) x Altre gestioni (altra pens.)	0.0029	Gestione (altra pens.): Altre gestioni	0.0056
(Recente = 1) x Ex-Inpdai (altra pens.)	-0.0027	Età del superst. alla liquidazione	-0.0054
(Femm. = 0) x Parasubordinati	0.0025	(Recent = 0) x Importo pens. al superst.	-0.0046
Anno decorrenza pens. dante causa	0.0025	(Femm. = 0) x Parasubordinati (altra pens.)	0.0043

Nota: (\*) *Recente* è una dummy che assume valore 1 se la pensione al superstite è stata liquidata dell'anno a cui si riferisce la campagna RED o nell'anno precedente. Se la pensione è stata liquidata prima assume valore 0.

Tabella 6 – La previsione delle pensioni con conguaglio a credito

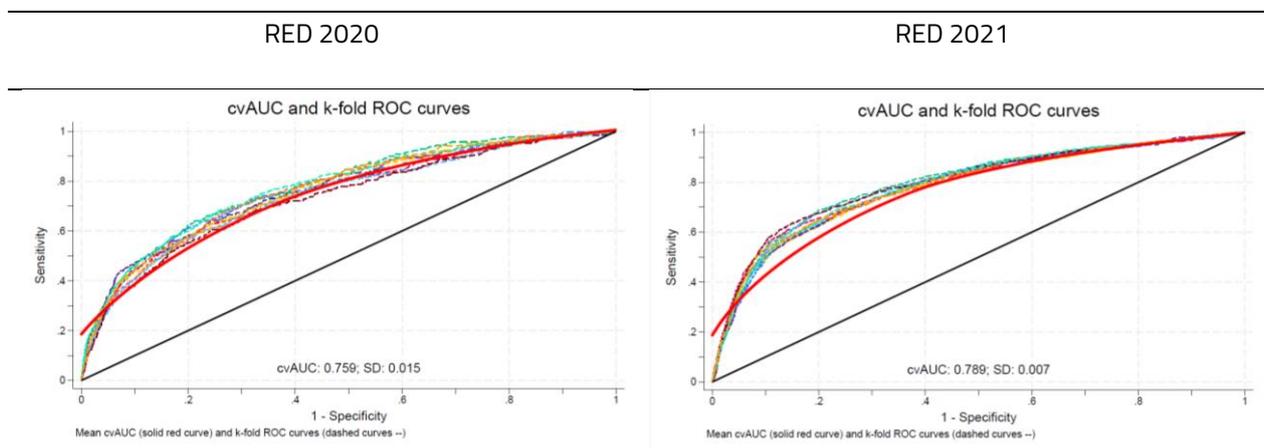
	2020	2021
Selected lambda	.0005634	.0003952
# of covariates	396	408
# of non-zero coefficients	149	244
(Recente* = 1) x Elettrici	0.0143	(Femm. = 1) x Anzianità contrib. altra pens. -0.0777
(Recente = 1) x Telefonici	0.0109	Anzianità contrib. (altra pens.) -0.0416
Importo pens. superst.	-0.0063	(Femm. = 1) x Importo altra pens. 0.0399
(Femm. = 1) x Elettrici	0.0056	Anno decorrenza altra pens. 0.0370
(Recente = 0) x Redd. lav. dipend. (pubbl.)	0.0055	(Femm. = 1) x Anno decorrenza altra pens. 0.0353
Età del superst. alla liquidazione	-0.0051	Età del superst. alla liquidazione 0.0290
(Femm. = 1) x Importo altra pens.	0.0051	(Recent = 0) 0.0287
(Femm. = 1) x Redd. lav. dipend. (priv.)	0.0046	Gestione: Trasporti 0.0225
(Recente = 1) x Altri Enti	0.0044	(Femm. = 0) x Importo pens. al superst. -0.0205
(Recente = 1) x Ex-dipendenti Inps	0.0043	(Femm. = 1) x FPLD (altra pens.): 0.0186
Gestione (altra pens.): ENPAM	0.0042	(Recent = 0) x Anzianità contrib. -0.0181
(Femm. = 0) x Anzianità contrib.	0.0036	(Femm. = 0) x Anzianità contrib. 0.0170
(Femm. = 1) x FPLD (altra pens.)	-0.0035	(Femm. = 1) x Altri enti (altra pens.): -0.0152
(Femm. = 1) x Commerc. (altra pens.)	0.0029	(Femm. = 1) x CDCM post 90 (altra pens.): 0.0150
Importo pens. al superst. al quadrato	0.0028	Importo pens. al superst. -0.0142
Gestione: Ex-dipendenti Inps	0.0028	(Recent = 0) x Trasporti 0.0124
Anno decorrenza (altra pens.)	0.0028	(Femm. = 0) -0.0121
(Femm. = 1) x Redd. lav. dip. (pubbl.)	0.0027	Provincia CT 0.0120
Provincia: BO	-0.0025	(Femm. = 1) x Reddito da lav. dip. (pubbl.) 0.0120
(Femm. = 0) x Artigiani (altra pens.)	0.0024	(Femm. = 0) x ENPAM (altra pens.) 0.0117
Anno decorrenza dante causa	0.0023	Gestione (altra pens.): Parasubordinati -0.0112
(Recente = 0) x Artigiani (altra pens.)	0.0020	(Femm. = 0) x Importo pens. superst. al quadr. 0.0102
(Recente = 1) x ENPAM	0.0020	(Recent = 0) x Ex-Inpdai 0.0084
Gestione: Telefonici	0.0020	(Femm. = 0) x Importo pens. dante causa 0.0076
(Femm. = 1) x Parasubord. (altra pens.)	0.0019	(Femm. = 0) x FPLD (altra pens.): -0.0071
Anni dalla liquidazione	0.0019	(Recent = 1) x Altri enti (altra pens.): -0.0067
Gestione: Commercianti	0.0018	(Recent = 0) x FPLD -0.0064
(Femm. = 0) x Importo altra pens.	-0.0018	Provincia NA 0.0061
(Recente = 0) x Ex-Inpdai	0.0018	Anni dalla liquidazione -0.0058
(Recente = 0) x Esodati (altra pens.)	0.0017	(Recent = 1) x real_imp_lordo_annoNS -0.0056

Nota: (\*) *Recente* è una dummy che assume valore 1 se la pensione al superstite è stata liquidata dell'anno a cui si riferisce la campagna RED o nell'anno precedente. Se la pensione è stata liquidata prima assume valore 0.

I Grafici 4 e 5 forniscono una misura della capacità previsiva del modello e quindi della sua capacità di individuare i superstiti a cui sono state liquidate pensioni di importo superiore o inferiore a quanto effettivamente dovuto. Come spiegato nel paragrafo precedente, tale misura coincide con la cosiddetta AUC (*area under the curve*). L'AUC del modello di previsione dei conguagli a debito basato sui dati della

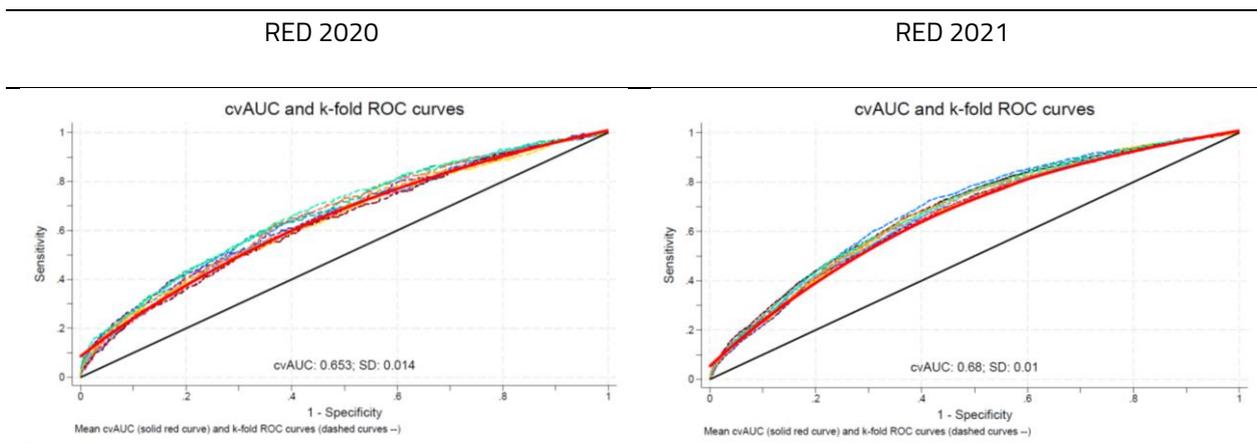
campagna RED 2020 (Grafico 4; pannello di sinistra) è di 75,9, con intervalli di confidenza al 95% di (75,07- 76.74), valori che indicano una capacità accettabile di discriminare tra pensioni al superstite definite correttamente e non-correttamente. L’AUC del modello stimato sulla base della RED 2021 (Grafico 4; pannello di destra) è di 78,9, con intervalli di confidenza al 95% di (64.38-66.14), un valore di poco sotto la soglia per cui la capacità di discriminare del modello possa essere definita eccellente. La capacità di discriminare le pensioni al superstite con conguagli a credito (Grafico 5) è leggermente minore. L’AUC basato sui dati della Campagna RED 2020 (pannello di sinistra) è di 65,3, con intervalli di confidenza al 95% di (64.49 – 66.24), mentre quello basato sulla RED 2021 (pannello di destra) è di 68,0, con intervalli di confidenza al 95% di (67.56 - 68.44), valori leggermente inferiori al valore soglia di 70 che denota una capacità di discriminare accettabile.

Grafico 4 – Capacità di identificare i trattamenti con conguagli a debito



Nota: Capacità di discriminare del modello misurata in termini di area media sotto la curva (AUC) calcolata mediante validazione incrociata a 10 campioni (10-fold cross-validation). La linea diagonale rappresenta un modello che discrimina casualmente (AUC = 50). L'asse delle ascisse mostra la proporzione di titolari di pensioni al superstite determinate correttamente, senza esiti di conguaglio a debito/credito, che il modello classifica erroneamente (falsi positivi). L'asse delle ordinate, invece, mostra la proporzione di titolari di prestazioni con conguagli a debito/credito che il modello classifica correttamente come tali (veri positivi). Ogni punto della curva (Receiver Operating Characteristic curve – ROC) rappresenta una coppia di percentuali di falsi positivi e di veri positivi al variare del valore soglia utilizzato per classificare le osservazioni in (1/0). CvAUC sta per “area media sotto la curva”, ed è calcolata mediante validazione incrociata. Pannello di sinistra (RED 2020): AUC=75.92; intervallo di confidenza al 95% corretto per bias mediante bootstrap: (75,07-76.74). Pannello di destra (RED 2021): AUC=78.87; intervallo di confidenza al 95% corretto per bias mediante bootstrap: (78.34-79.43).

## Grafico 5 – Capacità di identificare i trattamenti con conguagli a credito



Nota: Pannello di sinistra: AUC=65.31; intervallo di confidenza al 95% corretto per *bias* mediante *bootstrap*: 64.49 – 66.24. Pannello di destra: AUC= 67.99; intervallo di confidenza al 95% corretto per *bias* mediante *bootstrap*: 67.56 – 68.44.

Nel complesso, la metodologia proposta, basata su un modello probabilistico in cui la selezione delle covariate avviene in base al LASSO, si mostra promettente come strumento di *screening* per determinare, a seguito della verifica reddituale massiva, il rischio di erogare pensioni al superstite con importi superiori o inferiori rispetto alla misura effettivamente spettante. Il modello, infatti, fornisce una stima del rischio a livello di singolo titolare di pensione al superstite che l'importo del suo trattamento non sia corretto. Tali stime possono essere utilizzate dall'Istituto per identificare i superstiti per cui il rischio di conguagli è più elevato, e questo consente di prestare loro particolare attenzione già nella fase di definizione della pensione.

## 6. Sintesi e conclusioni

Dall'analisi dei dati ottenuti incrociando gli esiti delle campagne RED 2020 e 2021 con il Casellario centrale delle pensioni e gli archivi INPS dei rapporti di lavoro pubblici e privati. Le Campagne hanno generato conguagli a debito o credito dei titolari di pensione, con incidenze molto variabili tra gli anni. Nel 2020 i conguagli a debito erano circa il 4%, mentre nel 2021 sono saliti a quasi il 10% delle posizioni esaminate, rispettivamente l'1 e il 3% del totale delle prestazioni al superstite erogate. La mediana degli importi è però scesa da circa 400 euro l'anno a meno di 20 euro. Per quanto riguarda invece i conguagli a credito sono passati dal 4% delle posizioni esaminate nel 2020, a quasi il 20% nel 2021 e la media della distribuzione è intorno ai 60 euro per i conguagli rilevati nel 2021 e intorno ai 24 euro per quelli del 2021.

Nel complesso, i risultati dell'analisi supportano l'uso di modelli di previsione probabilistica LASSO per individuare le prestazioni con maggiore rischio di conguagli a debito/credito ex post e ridurre l'incidenza

ex ante. La metodologia proposta attinge a informazioni reperibili negli archivi INPS e stima il rischio di generazione di conguagli nella definizione della pensione a livello di singolo beneficiario. Tale rischio può essere utilizzato per distinguere tra i titolari di prestazioni ed identificare quelli per cui è più probabile la generazione di un conguaglio successivo alla verifica reddituale e adottare procedure che possano recepire tale metodologia. Si fa presente che si procederà ad un ulteriore fine-tuning sulla base dei dati della Campagna RED 2022 (anno reddito 2021) e del Casellario centrale delle pensioni del 2022.