

FACOLTÀ DI ECONOMIA LUISS Guido Carli

Dipartimento di Economia e Finanza

TESI DI LAUREA

In Matematica Finanziaria Progredita

***I MODELLI STOCASTICI: TEORIA E PRATICA
DELLE NUOVE TECNICHE PER LA
VALUTAZIONE DELLA RISERVA SINISTRI***

Relatore:

Prof. Gennaro Olivieri

Correlatore:

Prof. Marco Dall'Aglio

Candidato:

Cecilia Calcarella

ANNO ACCADEMICO 2011/2012

I MODELLI STOCASTICI: TEORIA E PRATICA DELLE NUOVE TECNICHE

PER LA VALUTAZIONE DELLA RISERVA SINISTRI

Indice

1. INTRODUZIONE

1.1 QUESTIONI PRELIMINARI

1.1.a COSA

1.1.b PERCHE'

1.1.c CHI

Normativa Nazionale

Solvency II

I Pilastro

II Pilastro

III Pilastro

1.1.d COME

SEZIONE I: TEORIA

CAPITOLO I

2. MODELLI STOCASTICI

2.1 PREMESSA: MODELLI STOCASTICI E MODELLI DETERMINISTICI

2.2 DEFINIZIONI FONDAMENTALI

Le Coorti dei Sinistri: i Dati

Il Triangolo dei Run-Off

2.3 GENERALIZED LINEAR MODELS

2.4 IL METODO DELLA CATENA O CHAIN-LADDER

Storia e Premesse

L'ipotesi fondamentale ed il funzionamento del Chain-Ladder

L'inflazione: alcuni aggiustamenti

Metodo di separazione: cenni

2.5 MODELLI CHE REPLICANO LA STRUTTURA DEL CHAIN-LADDER

2.5.a. OVER-DISPersed POISSON MODEL (ODP)

Stima dei Parametri del Modello: La Funzione di quasi-Verosimiglianza

Calcolo delle Riserve e del Prediction Error (RMSEP)

Una proposta alternativa: Verrall, Stoccolma 2007

Concludendo: scopi e vantaggi dell'ODP

2.5.b MODELLO 'NEGATIVE BINOMIAL'

Derivazione del modello Negative Binomial dall'ODP

Stima dei Parametri: funzione di Verosimiglianza

Calcolo delle Riserve, Process Variance ed Estimation Variance

Riepilogo delle Formule

Limite del modello e la sua Approssimazione Normale

2.5.c 'NORMAL APPROXIMATION TO NEGATIVE BINOMIAL'

La procedura joint modelling: due GLM per la stima della dispersione

Calcolo delle Riserve, Process Variance ed Estimation Variance

2.5.d IL MODELLO DI MACK

Descrizione del Modello

Costi Ultimi: Best Estimate e MSEP

Confronto tra metodi simili: ODP, DFCL e NANB

2.6 MODELLI CON ALTRE DISTRIBUZIONI

Il Modello a Distribuzione Log-Normale

Stima del Prediction Error del Modello

Vantaggi e Limiti del Modello

Distribuzione Gamma

2.7 MODELLI A STRUTTURA DIFFERENTE DAL CHAIN-LADDER

2.7.a LA CURVA DI HOERL

2.7.b MODELLO DI WRIGHT

2.8 GAM E SMOOTHING MODELS

Smoothers: Premesse

Dagli Smoothers ai Modelli Additivi Generalizzati

Smoothers nella riserva sinistri

2.9 I MODELLI STOCASTICI E SOLVENCY II

2.10 CONFRONTI TRA MODELLI

CONCLUSIONI

CAPITOLO II

3. IL BOOTSTRAPPING ED ALTRI METODI DI SIMULAZIONE

3.1 INTRODUZIONE

3.2 LA TECNICA DI BORNHUETTER-FERGUSON

I limiti della struttura Chain-Ladder

La Tecnica BF

3.3 IMPARARE DALL'ESPERIENZA: LA STATISTICA BAYESIANA

Il Teorema di Bayes

Prior/Posterior Predictive Distributions

Vague Priors: distribuzioni non informative

Limiti computazionali

3.4 MARKOV-CHAIN MONTE-CARLO

3.5 METODI BAYESIANI

3.6 BOOTSTRAPPING

Metodi analitici e tecniche di simulazione

Breve storia del Bootstrap

L'idea di base: "Re-sampling with replacement"

Bootstrapping nella stima della riserva sinistri

I residui (di Pearson)

Pro e contro

Questioni generali

3.7 TECNICHE DI SIMULAZIONE ALTERNATIVE AL BOOTSTRAPPING

Simulazione dai parametri

I metodi bayesiani

CONCLUSIONI

SEZIONE II: PRATICA

CAPITOLO III

4. I MODELLI E LA LORO APPLICAZIONE ALLA VALUTAZIONE DELLA RISERVA SINISTRI

4.1 INTRODUZIONE

4.2 LOSS RESERVING WITH SOFTWARES

4.2.a R Software

Il pacchetto ChainLadder

4.2.b Implementazione dei modelli

I Dati

Il Chain-Ladder deterministico

GLMs: Il modello Over-Dispersed Poisson

Il modello di Mack

Il modello Log-Normale

Il modello Gamma

Bootstrap

CONCLUSIONI

APPENDICE – Script dei comandi utilizzati in R

5. CONCLUSIONI

BIBLIOGRAFIA

Ringraziamenti

1. INTRODUZIONE

Le imprese di assicurazione sono enti peculiari nel panorama dei soggetti economici e degli intermediari finanziari, in quanto caratterizzate da diversi insoliti aspetti. Un ciclo economico invertito, la lunghezza dei processi assicurativi e di investimento-disinvestimento dei fondi, ed altre caratteristiche contraddistinguono la categoria. Quanto descritto è da ricondursi all'attività tipicamente svolta: ciò di cui le imprese di assicurazione si occupano è la vendita di promesse.

Ad ogni modo, è sin da subito bene effettuare una precisazione: assicurarsi contro un determinato evento non significa evitare il suo verificarsi, bensì proteggersi dagli effetti negativi ad esso relativi. Ciò si traduce nella funzione e nella finalità dell'assicuratore, rintracciabili proprio nell'offerta di garanzie patrimoniali per gli assicurati, tramite prestazioni di carattere *futuro ed eventuale*.

In generale dunque, la solvibilità implica la disponibilità di mezzi finanziari sufficienti a soddisfare gli impegni presenti e futuri nei confronti degli assicurati su un orizzonte temporale di riferimento. Ciò avviene con un prestabilito livello di confidenza, corrispondente alla probabilità del percentile della distribuzione della variabile aleatoria "valore complessivo del danno", assunto come target di sicurezza.

La possibilità di sopravvivenza è esaminabile principalmente secondo due prospettive. La prima si sofferma sull'esposizione complessiva dell'impresa *alle fonti di rischio* (di natura tecnica, finanziaria, o semplicemente generiche situazioni di rischio). La seconda guarda invece alla disponibilità di un *capitale di sicurezza*, che possa essere sufficiente a fronteggiare perdite inattese, e superare possibili crisi.

Fondamentale è porre l'attenzione su come l'inversione del ciclo produttivo, per cui i ricavi precedono i costi, comporti maggiori elementi di rischiosità rispetto alle altre imprese industriali. In proposito, le Compagnie di Assicurazione utilizzano opportune metodologie statistiche ed attuariali, al fine di tutelarsi dai futuri esborsi dovuti a sinistri ovvero costi di gestione, costituendo così le cosiddette *riserve tecniche*. E sono proprio le riserve tecniche la prima misura per salvaguardare la solvibilità delle imprese di

assicurazione, ben prima del margine di sicurezza: un costante controllo ed una valutazione il più possibile corretta delle riserve sono senza dubbio di primaria importanza.

Nel contesto metodologico descritto, e muovendo dal contributo originale di England e Verrall (*British Actuarial Journal*, 8, 2002), ci si propone di presentare differenti metodologie, accomunate dallo scopo di ottenere la miglior stima delle riserve e degli accantonamenti necessari. I classici modelli deterministici si limitano a fornire un valore puntuale degli impegni di risarcimento aleatori; il ricorso a metodi stocastici consente invece di ottenere, in aggiunta a tale posta, intervalli di variazione della stessa, secondo prefissati livelli di probabilità. Specialmente, fanno ciò in armonia con l'impianto metodologico richiesto dalle più recenti normative e direttive, quali le proposte del progetto Solvency II, presentato più avanti. Nella medesima ottica si illustra la tecnica del bootstrap, e le tecniche basate sulla simulazione dei dati.

Il lavoro è stato diviso in due macro-sezioni, differenti eppure complementari. La prima, un'impostazione teorica, parte dalla descrizione dei concetti fondamentali legati alle riserve sinistri nelle compagnie di assicurazione. La sezione introduttiva riguarda perciò aspetti generali; è stata pensata per presentare al lettore la questione rispondendo con ordine ai classici "*Cosa? Perché? Chi? Come?*".

Nel primo Capitolo vero e proprio sono stati osservati più da vicino i modelli stocastici. In primis, è condotta una rapida presentazione della tecnica chain-ladder, tradizionalmente deterministica ma sempre punto di riferimento e di partenza. Sono quindi descritti i modelli a carattere stocastico che ne replicano la struttura (il modello Over-Dispersed Poisson ed il distribution-free di Mack tra tutti), poi le curve parametriche, infine i modelli ricompresi nel framework dei Generalized Additive Models.

Nel Capitolo II si compie un ulteriore passo: se i metodi analitici del primo restano legati alle difficoltà matematiche, la possibilità di replicare la realtà *simulandola* consente da un lato l'abbandono delle problematiche di calcolo, dall'altro di ricavare preziose informazioni aggiuntive. Seguendo un percorso logico che parte da nozioni di Statistica Bayesiana, incontra i metodi MCMC e si sofferma sulla tecnica di Bornhuetter e Ferguson, viene introdotta l'innovativa metodologia del bootstrapping. L'importante risultato che in tal maniera si ottiene è *una* distribuzione predittiva completa, a partir

dalla quale derivare poi ogni misura d'interesse desiderata. Si tratta dunque di una tecnica sia più semplice, concettualmente ed a livello di implementazione informatica, sia potente e completa.

In ultimo, la seconda sezione viene a coincidere con il terzo Capitolo, probabilmente il cuore del lavoro: l'implementazione pratica dei modelli, attraverso il costante riferimento ad Excel oltre che – in modo particolare – al software statistico R. Si sottolinea il sorprendente dato sull'uso dei modelli stocastici e simulativi: esso risulta ancora piuttosto scarso, nonostante i recenti sviluppi in ambito internazionale. Indagini condotte hanno dato voce agli utilizzatori dei modelli, facendo emergere di fatto la comune difficoltà nel reperimento ed uso di software informatici adatti. Si è quindi inteso presentare un utilizzo "pratico" degli stessi modelli descritti in "teoria" nei capitoli precedenti: con costanti richiami alla sezione teorica, e con il supporto di R (ampliato da opportuni pacchetti aggiuntivi), sono illustrati i passaggi, i comandi, ed i principali risultati di una valutazione della riserva sinistri per una Compagnia operante nel ramo danni.

1.1 QUESTIONI PRELIMINARI: COSA (RISERVE) PERCHE' (RISCHIO) CHI (NORMATIVA E SOLVENCY II) COME (MODELLI)

1.1.a COSA

Tutte le imprese autorizzate ad esercitare l'attività di assicurazione sono chiamate ogni anno a redigere il bilancio di fine esercizio. Tenendo conto di quanto affermato, è indiscussa l'importanza ricoperta dalla riserva tecnica, terza delle passività dello stato patrimoniale. L'impresa è infatti obbligata alla costituzione di riserve tecniche, secondo l'articolo 23 del decreto legislativo 175/1995¹.

Con il termine riserve tecniche si intende l'accantonamento necessario che l'impresa deve disporre per far fronte agli impegni assunti verso gli assicurati, ed in quanto tali sono determinate - prevalentemente - sommando alla riserva premi la riserva sinistri. Per entrambe, in sede di bilancio, la consistenza è da valutarsi con la massima accuratezza consentita, dai dati e dai modelli a disposizione.

La riserva premi, definita anche *riserva per rischi in corso o per danni non ancora denunciati*, andrebbe calcolata per ogni contratto singolarmente preso, seguendo il criterio cosiddetto *pro rata temporis*. Il calcolo, fondato sui premi di tariffa (al netto delle le spese di acquisizione e le imposte a carico degli assicurati), è semplice e facilmente effettuabile, seppur accompagnato da costi. Diverso è il discorso per quanto concerne la riserva sinistri (in inglese *claims reserving*), indice dell'impegno futuro della compagnia verso assicurati o danneggiati. La valutazione di questa sottovoce richiede un impegno nettamente superiore. Si rende necessario l'utilizzo di molte risorse, in particolare l'elaborazione dell'informazione riportata dall'esperienza del passato. Fondamentale è la formulazione di *previsioni* su grandezze economiche tutt'altro che statiche, bensì evolventisi in condizioni di incertezza. Si pensi, ad esempio, all'aleatorietà dell'inflazione economica generale o della classe di rischio, alle sentenze della magistratura, in genere tendenzialmente

¹ *Articolo 23, Riserve tecniche relative al portafoglio italiano, comma 1: "Le imprese hanno l'obbligo di costituire, per i contratti del portafoglio italiano, riserve tecniche sufficienti a garantire le obbligazioni assunte. Dette riserve debbono essere costituite al lordo delle cessioni in riassicurazione ed in conformità agli articoli 31,32,33,35, 36 e 37, commi 1 e 2, del decreto legislativo di attuazione della Direttiva 91/674 CEE del 19 dicembre 1991".*

TEORIA - CAPITOLO I
2. MODELLI STOCASTICI

“It is sometimes hoped that stochastic methods will provide solutions to problems when deterministic methods fail. This rather misses the point. The usefulness of stochastic models is that they can provide more information, which may be useful in the reserving process and in the overall management of the company”
P. D. England and R. J. Verrall

2.1 PREMESSA: MODELLI STOCASTICI E MODELLI DETERMINISTICI

La solvibilità cui aspira Solvency II non è quella statica - o almeno non solo - bensì quella dinamica, che sia in linea con i repentini cambiamenti del mercato e sottenda la naturale dinamicità dell'attività dell'impresa. È per raggiungere questo particolare tipo solvibilità che si rende, di fatto, fondamentale basare le valutazioni delle grandezze rilevanti su modelli di tipo *stocastico*, più che *deterministico*.

Prima di passare in rassegna alcuni dei principali modelli stocastici attualmente in uso nella pratica assicurativa del ramo danni, ci si sofferma di seguito sulle principali differenze tra queste due macrocategorie. L'impiego di metodologie di calcolo di uno o dell'altro tipo per la valutazione della riserva sinistri ricopre il ruolo di uno degli argomenti più dibattuti negli ultimi anni, nella letteratura attuariale riferita alle assicurazioni contro i danni.

Le tecniche deterministiche, che costituiscono le più classiche e tradizionali, presentano l'indiscusso vantaggio di essere rapidamente implementabili sotto il profilo computazionale. Minori vincoli sui dati di base conferiscono un certo grado di semplicità d'uso; soprattutto, risultano di immediata comprensione a livello manageriale di impresa. Per tale motivo gli attuari, sia in Italia che all'estero, hanno in genere preferito i metodi deterministici rispetto a quelli stocastici.

Tra i punti di debolezza, invece, figura il fatto che queste valutazioni restituiscano soltanto un valore puntuale della previsione, trascurando ogni possibile valutazione sulla probabilità del suo realizzarsi. In altri termini, possono sì da un lato produrre una stima “migliore” (best estimate) nel ventaglio dei risultati possibili, ma dall'altro non sono in grado di misurare la bontà di questa stima.

Non solo: i modelli deterministici sono spesso utilizzati senza una chiara conoscenza delle assunzioni che si stanno ponendo, cosa che non accade con i modelli stocastici, in quanto basati su ipotesi del tutto esplicite ed verificabili con una grande varietà di tecniche.

Ad ogni modo, il quadro normativo che si sta delineando richiede sempre di più la diffusione di metodologie stocastiche. Tra i vari aspetti, infatti, il CEIOPS ha posto come obiettivo il raggiungimento di un benchmark armonizzato in chiave probabilistica quale requisito regolamentare per la riserva sinistri. Si tratta di un obiettivo che può essere raggiunto solamente mediante l'utilizzo di un modello di valutazione di tipo stocastico alla serie storica dei sinistri.

Soltanto i modelli stocastici, nonostante siano evidentemente caratterizzati da una maggiore complessità applicativa, e da ipotesi più impegnative sui dati di base, permettono di conseguire, oltre alla stima puntuale della riserva, intervalli di variazione della stessa secondo un prefissato livello di probabilità. Di fatto essi sono atti a fornire una misura di precisione della BE, trattando il processo che porta alla determinazione della riserva come un esercizio di analisi dei dati su cui basarsi. In altri termini: sono in grado di venir incontro e soddisfare il crescente interesse che si sta sempre più esprimendo in ambito internazionale verso la stima di un *downside potential*. Un'altra importante questione è proprio quella della stima: la BE è semplicemente un valore di tipo *puntuale* dei futuri esborsi. Al contrario, è evidente che i pagamenti reali potranno differire da quelli previsti e, mentre i modelli deterministici non danno un'idea di quanto possa essere significativo questo scostamento, i metodi stocastici forniscono invece un intervallo entro i cui limiti ci si può attendere cada l'importo dei futuri esborsi con un determinato livello di confidenza.

Anche i metodi stocastici presentano alcuni svantaggi. Si pensi, ad esempio, al fatto che essi modellano una serie di eventi molto ampia attraverso un numero relativamente esiguo di parametri. Un'ulteriore critica riguarda le assunzioni, definite troppo semplicistiche ed a volte irreali, dovute anche alla rigidità di questi modelli. Questi, in effetti, che non consentono facilmente di incorporare giudizi o fattori esterni.

Inoltre, anche l'applicazione pratica e computazionale risulta frequentemente complessa da attuare richiedendo un'abilità statistica e di calcolo non indifferente.

Nonostante ciò i modelli stocastici risultano essere efficaci: l'utilità che maggiormente ci interessa è quella derivante dalla maggiore informazione fornita, il che può risultare utile alla compagnia non solo nel processo di costituzione della riserva, ma anche nella sua complessiva gestione.

In ultimo, un importante aspetto ed obiettivo su cui si va a concentrarsi in questo Capitolo sarà il concetto del *fitting-of-the-model*: un metodo stocastico per la stima della riserva sinistri prevede, dopo la scelta di una struttura parametrica appropriata, l'accostamento di tale struttura ai dati disponibili (nel caso in esame: al triangolo superiore dei run-off, noto). L'approccio tipico è quella della stima dei parametri della struttura, vera ma incognita, con il metodo della massima verosimiglianza.

2.4 IL METODO DELLA CATENA O CHAIN-LADDER

Storia e Premesse

Il metodo della catena, conosciuto anche come *chain-ladder method*, è probabilmente il più diffuso nella pratica per la stima della riserva sinistri. Le origini del metodo rimangono sconosciute: non si è a conoscenza di chi ne sia stato l'autore o chi l'abbia applicato per la prima volta al triangolo di run-off. Taylor afferma che già agli inizi degli anni Settanta, il professore R. Beard, consulente del Department of Trade del Regno Unito, faceva ampio ricorso a tale metodo. Inoltre, gli autori inglesi Clarke e Harland vi si riferiscono in un testo del 1974: la nascita del metodo è pertanto da porre in data antecedente. Addirittura, pare che esso appaia già nel 1973 come applicazione della Teoria della Credibilità alla riserva sinistri.

Il chain-ladder ricade nella categoria definita "*fenomenologica*" da Taylor, McGuire e Greenfield (2003)². Questo poiché non riflette i meccanismi sottostanti al pagamento degli indennizzi, e di conseguenza la necessità di *una forma di correzione potrebbe non essere evidente*.

La denominazione del metodo in esame, "*metodo chain-ladder*" o "*metodo della catena*", è direttamente imputabile ai meccanismi di calcolo e di stima. Alla base delle stime c'è la proiezione al futuro dell'esperienza passata in materia di liquidazione, supponendo che l'assicuratore non muti nel tempo la modalità di chiusura dei sinistri. In questo modo è chiaro come i valori futuri dipenderanno fortemente dai valori passati definendo in tal maniera una struttura concatenata. Appare chiaro, al contempo, un limite del modello: tendenzialmente rigido, difficilmente adattabile ai mutamenti esterni.

Ancora, si tratta di una tecnica d'applicazione estremamente semplice, e non necessita di alcuna ipotesi statistica riguardo la distribuzione dei dati di partenza. Purtroppo, si vedrà meglio a breve, il chain-ladder si rivela decisamente sensibile alle variazioni di questi ultimi, correndo il rischio che un suo uso meccanico risulti non allineato ai principi di Solvency II, e dalla più recente evoluzione normativa tutta.

²Nella classificazione dei Loss Reserving Model secondo una struttura algebrica, Taylor, McGuire e Greenfield presentano uno spettro di possibilità che va da un estremo fenomenologico ad uno micro - strutturale. Nel primo si includono i modelli che descrivono l'esperienza nel pagamento di indennizzi, senza poggiare su alcun significato fisico diretto. I modelli che invece incorporano la struttura dello stesso processo di pagamento sono, appunto, i micro-strutturali.

2.5.a OVER-DISPERSED POISSON MODEL (ODP)

Il modello di Poisson con sovra-dispersione (a struttura chain-ladder nella presente analisi) differisce dal tradizionale modello di Poisson per la sua varianza. Quest'ultima, infatti, è proporzionale alla media, maggior o minore sulla base del valore di φ , il cosiddetto *parametro di dispersione*. Vigge sempre l'ipotesi di indipendenza tra i pagamenti incrementali.

Si ricorre alla variante sovra-dispersa poiché il modello di Poisson semplice, di cui pure si fa largo uso in statistica, non si dimostra qui pienamente efficace nel modellizzare l'andamento dei risarcimenti.

England e Verrall (2002) applicano il modello ODP direttamente ai sinistri incrementali

$$C_{ij} \sim iid ODP(m_{ij}).$$

Questi si distribuiscono con media e varianza, rispettivamente pari a

$$E(C_{ij}) = m_{ij} = x_i y_j$$

$$Var(C_{ij}) = \varphi E(C_{ij}) = \varphi x_i y_j$$

Dove:

- φ è un parametro incognito la cui stima parte dai dati stessi;
- x_i rappresenta il risarcimento atteso complessivo pagato per sinistri generati dall' i -esima generazione³;
- y_j la proporzione di tale ammontare pagata nel j -esimo anno di sviluppo. Vale pertanto:

$$\sum_{k=1}^n y_k = 1.$$

³ x_i è detta "expected ultimate claims", dove con il termine 'ultimate' si vuole intendere il valore che si ha nell'ultimo anno di antidurata osservato nel triangolo dei run-off.

2.5.c 'NORMAL APPROXIMATION TO NEGATIVE BINOMIAL'

Un approccio simile a quanto visto nello scorso paragrafo conduce con naturalezza alla definizione di un modello in cui

$$\begin{aligned} E(C_{ij}) &= (\lambda_j - 1)D_{i,j-1} & \text{Var}(C_{ij}) &= \varphi_j D_{i,j-1} \\ E(D_{ij}) &= \lambda_j D_{i,j-1} & \text{Var}(D_{ij}) &= \varphi_j D_{i,j-1} \end{aligned}$$

L'assunzione sui dati, sia incrementali che cumulati, è che siano approssimativamente distribuiti secondo una Normale. Rispetto al modello precedente, la media è inalterata, tuttavia la varianza risulta aggiustata in modo tale da rendere possibile $\lambda_j < 1$. Ovviamente il modello è valido qualora $D_{i,j-1}$ sia noto, e si tratta ancora di un metodo ricorsivo. In generale, valgono le stesse osservazioni fatte per il *Negative Binomial*.

È importante sottolineare che nel modello '*Normal Approximation to Negative Binomial*' il parametro di dispersione è funzione dell'antidurata/anno di sviluppo. Nelle formule si fa riferimento a φ_j : il parametro con indice j . Per ricavarne un valore appropriato England e Verrall propongono una procedura 'joint modelling'. Si tratta di un meccanismo iterativo, che richiede la stima di parametri per la media e la varianza assieme, piuttosto che per la media solamente. I dettagli sono rintracciabili nella discussione sui GLM di McCullagh e Nelder (1989) e in Renshaw (1994).

I dati vengono normalmente considerati in forma cumulata. Chiamando $D_{i,j-1} = w_{ij}$

$$\begin{aligned} E(D_{ij}) &= \lambda_j D_{i,j-1} & \text{Var}(D_{ij}) &= \varphi_j D_{i,j-1} \end{aligned}$$

e definendo

$$f_{ij} = \frac{D_{ij}}{D_{i,j-1}} = \frac{D_{ij}}{w_{ij}}$$

si ricava

$$\begin{aligned} E\left(\frac{D_{ij}}{w_{ij}}\right) &= \lambda_j & \text{Var}\left(\frac{D_{ij}}{w_{ij}}\right) &= \frac{\varphi_j}{w_{ij}} \\ E(f_{ij}) &= \lambda_j & \text{Var}(f_{ij}) &= \frac{\varphi_j}{w_{ij}} \end{aligned}$$

2.9 I MODELLI STOCASTICI E SOLVENCY II

A questo punto dello studio, si intende spiegare come i modelli stocastici possano rispondere alle esigenze più attuali, e cercare capire l'ispirazione al progetto Solvency II dietro essi.

Vediamo brevemente le principali differenze fra l'approccio tradizionale del margine di solvibilità (sostanzialmente basato su un'impostazione contabile e su parametri dimensionali più che una valutazione del rischio) e l'approccio di Solvency II, fondato invece su una valutazione della prospettiva economica. I modelli sul margine di solvibilità hanno un riferimento temporale molto ristretto, mentre Solvency II abbraccia un'ottica di lungo termine: sia gli attivi che i passivi sono valutati mark-to-market. Soprattutto, *la valutazione stocastica del rischi non era considerata nel metodo tradizionale, mentre è fondamentale per Solvency II.*

Stando a quanto disposto da Solvency II, le imprese di assicurazione possono utilizzare *modelli interni* per il calcolo di uno o più sottomoduli di rischio. Previa autorizzazione dall'Autorità di Vigilanza, è possibile sostituire il capitale di sicurezza (SCR, cfr. Sezione Introduttiva) calcolato dalla formula standard con quello risultante dall'applicazione di un modello implementato all'interno dell'impresa. Una simile possibilità ha lo scopo di determinare un requisito di capitale che sia specchio del profilo di rischio dell'impresa, e che incentivi il miglioramento della gestione interna dei rischi.

In altri termini, questa impostazione prima richiede di "costruire" la distribuzione di probabilità della variabile rappresentativa del valore d'impresa (tra un anno), successivamente collega l'SCR a questa distribuzione di probabilità.

Solvency II domanda alle imprese la determinazione di più SCR, uno per ognuna delle fonti di rischio. La valutazione può essere effettuata con la metodologia semplificata e standard, oppure tramite il ricorso ad un modello interno. Tale approccio serve a descrivere la dinamica stocastica dei fattori di rischio che caratterizzano l'impresa, nel modo più possibile preciso. Alla luce dello stato attuale delle cose, sappiamo che il nuovo sistema di calcolo delle riserve dovrebbe basarsi sulla somma delle due già viste componenti:

CONCLUSIONI

I temi appena trattati sono oggetto di dibattito assicurativo ormai da molti anni, così come appare nella recente letteratura attuariale. L'avvento del Solvency II rappresenta un'ulteriore spinta nella direzione di stabilità, sicurezza, prudenza e trasparenza, e metodi che siano in grado di assicurare tutto ciò. La questione posta dal limite derivante da una stima puntuale si supera con i modelli stocastici; essi fanno assunzioni riguardanti la variabilità dei futuri pagamenti oltre che il valore atteso dei medesimi. Questi sono pertanto gli unici in grado di rispondere alle esigenze dei progetti internazionali, e sono perciò raccomandati nella determinazione del capitale di rischio. In particolare, i modelli lineari generalizzati sono già stati ampiamente impiegati nella valutazione delle riserve.

Nel presente Capitolo si è principalmente inteso offrire una descrizione dei principali modelli in questa categoria. L'approccio statistico generale consiste nella preliminare specificazione di un modello; successivamente si individua una stima delle richieste e dei dati sottostanti il modello selezionato (tipicamente ciò viene fatto tramite procedimenti di massima verosimiglianza). Al termine del processo, il modello può essere utilizzato per costruire stime e statistiche, e valutarne la precisione.

Punto di partenza, il più classico ed utilizzato metodo della catena. La tecnica del chain-ladder, si è visto, usa come misura di variabilità il *Prediction Error*, la deviazione standard della distribuzione delle possibili riserve. Sarebbe consigliabile tener conto di ulteriori fattori, come ad esempio la possibilità di imprevisti – fatto, questo, che aumenterebbe l'incertezza: tuttavia, ciò complicherebbe eccessivamente la definizione del modello.

Il campo d'analisi è stato poi allargato, estendendolo ai modelli stocastici: a struttura chain-ladder e diversa dal chain-ladder. Tra i primi, il modello Over-Dispersed Poisson, quello *Negative Binomial* e la sua approssimazione Normale, ed il modello proposto da Mack, che non specifica alcuna distribuzione iniziale per i dati. Si tratta di metodi che replicano quasi perfettamente i risultati ottenibili con la tecnica tradizionale; in tutti questi casi, l'espressione del predittore lineare ha sempre mantenuto la tipica impostazione, notevolmente parametrizzata:

$$\eta_{ij} = c + \alpha_i + \beta_j.$$

È stato in particolar modo per sopperire ad una tale parametrizzazione che modelli alternativi, caratterizzati da una struttura ben diversa, sono stati introdotti nella discussione: tra questi, la Curva Parametrica di Hoerl ed il modello suggerito da Wright. Il predittore lineare, in tal caso, è specificato dall'equazione:

$$\eta_{ij} = c + \alpha_i + \beta_i \log(j) + \gamma_j.$$

Tra i grandi vantaggi di questi metodi non parametrici sta la possibilità di estrapolazione.

Infine, modelli ispirati non più ai GLMs, bensì ai Generalized Additive Models, sono stati descritti, tramite la generalizzazione del predittore precedente:

$$\eta_{ij} = s_{\theta_i}(i) + s_{\theta_j}(\log(j)) + s_{\theta_j}(j).$$

resa possibile dalle funzioni spline e dai parametri di smoothing.

Per ogni modello è stato possibile non solo effettuare una stima delle riserve necessarie a far fronte agli impegni futuri, ma anche procedere con un calcolo – di volta in volta impostato ben diversamente – della bontà della stima. Per ogni metodo si è fatto riferimento al MSEP, come somma di *Estimation* e *Process Variance*. Idealmente, ottimale sarebbe riuscire nel tracciato di una completa distribuzione predittiva di probabilità, a partire dalla quale sarebbe possibile l'estrazione di percentili ed altre statistiche utili.

Risulta sempre importante poter disporre di una base informativa sufficientemente ampia ed attendibile. Va sottolineato che l'impiego di un metodo è condizionato dal tipo e dall'entità dell'informazione disponibile; la scelta poi di uno tra i metodi possibili è influenzata dalle particolarità del portafoglio: non esiste un modello adatto ad ogni situazione, in modo standard.

A questo punto si rende necessaria una precisazione: può accadere talvolta che si speri o ci si aspetti ingenuamente che un metodo stocastico sia in grado di offrire una soluzione laddove un metodo più tradizionale ha precedentemente fallito. Questo non è purtroppo corretto: l'utilità più profonda delle tecniche stocastiche sta nel maggior

numero di informazioni che sono in grado di fornire, rispetto alle deterministiche. I modelli stocastici sono la chiave di volta per accedere alla variabilità della riserva sinistri, darne definizione e tenerne conto secondo il modo migliore e più equilibrato possibile.

A partire da quanto detto, si vedrà più avanti come sia possibile implementare praticamente diversi dei modelli su supporti informatici. Ad esempio, nel caso dei modelli a struttura a catena, i passaggi fondamentali saranno ridotti all' inserimento dei dati, la creazione dei parametri con un arbitrario valore iniziale, il calcolo del predittore lineare e della media, e successivamente della funzione di log-verosimiglianza per ogni cella del triangolo, infine, la loro somma e massimizzazione.

Quotando England e Verrall (2002): *“it has to be borne in mind that traditional techniques were developed before the advent of desktop computers, using methods which could be evaluated using pencil and paper. With the continuous increase in computer power, it has to be questioned whether it would not be better to examine individual claims rather than use aggregate data”*.

TEORIA - CAPITOLO II
3. IL BOOTSTRAPPING
ED ALTRI METODI DI SIMULAZIONE

“*We can do it the easy way, or we can do it the hard way*”, P.D. England

3.1 INTRODUZIONE

La valutazione stocastica delle riserve sinistri è negli ultimi anni divenuta un argomento sempre più importante, al punto di ispirare corsi accademici, articoli, presentazioni e conferenze in ambito attuariale, e molto altro. Una volta scelto il modello stocastico da impostare, la variabilità della riserva sinistri può essere ottenuta seguendo due differenti sentieri. Quello che England chiama “*the hard way*” è il metodo classico, quello analitico. Con “*doing it the easy way*”, England si riferisce invece ai metodi basati sulla simulazione dei dati.

Gli stimatori delle riserve sono spesso funzione delle osservazioni, e come si è visto possono essere anche analiticamente complessi. Pertanto, potrebbe essere talvolta difficile derivare le corrette espressioni analitiche.

Entra qui in gioco il cosiddetto *bootstrapping*, incentrato sul campionamento con ripetizione. Esso ha visto la sua popolarità incrementare appena introdotto per il chain-ladder, da England e Verrall (1999) ed England (2002). Ad ogni modo, dal momento che le tecniche esistenti di bootstrap adottano le ipotesi statistiche in letteratura, esse sono state pensate e costruite per offrire una misura della precisione della *best estimate* dell’attuario, senza tuttavia la possibilità di apportare modifiche alla stima in sé. Il vero obiettivo del bootstrapping risiede nell’ottenere una distribuzione di probabilità per il campione dei dati.

Il bootstrapping ha dimostrato nel tempo la sua grande utilità come tecnica statistica, ed ha destato più volte interesse; alcuni esempi:

- nel 1979 viene introdotto come metodo di calcolo per lo standard error del parametro di interesse di cui si intende ottenere stima (sempre nel caso si disponga di un campione di dati, provenienti da non nota distribuzione di probabilità);

- nel 1986 F. Ashe lo applica ai modelli PPCI e PPCF, ricorrendo anche al concetto di GLMs;
- nel 1988 e nel 1993 rispettivamente, G. Taylor e Brickman affrontano discussioni sull'argomento;
- nel 1996, R. Verrall implementa il bootstrapping per il GAMs ;
- nel 1999 P. England e R. Verrall usano il bootstrapping per la tecnica chain-ladder, usando i GLMs, e successivamente pubblicano l'articolo "*Stochastic Claims Reserving in General Insurance*", occupandosi di bootstrapping, dei modelli c.d. MCMC, e cercando di ricavare una distribuzione predittiva per le riserve sinistri;
- sempre England e Verrall pubblicano nel 2006 il paper "*Predictive Distributions – Bootstrap and Bayesian models (MCMC): ODP with varying scale, negative binomial, and Mack's model*";
- tra il 2007 e il 2008, Gibson e Bruce criticano il bootstrapping;

senza contare i numerosi articoli apparsi negli anni, tra cui si ricordano Pinheiro, Bjorwall, eccetera.

Si parta da una premessa: il bootstrap può essere applicato per qualsiasi modello univocamente descrivibile. I passaggi fondamentali, più avanti meglio descritti, vanno dalla definizione e *fitting* del modello stocastico, alla derivazione dei residui e di "pseudo-dati", ad un riadattamento del modello ed infine il calcolo di previsioni.

Ma subito prima di concentrare l'attenzione sull'efficacia – ma al tempo stesso sulla semplicità – del bootstrap, alcune questioni devono esser chiarificate al lettore: modelli MCMC, tecniche di Bayes e Bornhuetter-Ferguson.

3.5 METODI BAYESIANI

L'approccio bayesiano per il calcolo della stima della riserva sinistri ricomprende metodologie che prevedono la simulazione da distruzioni di probabilità ipotizzate a priori, poi aggiornate sulla base dell'esperienza sfruttando il teorema di Bayes e tecniche Monte Carlo Markov Chain.

Un ostacolo da non sottovalutare è rappresentato dalla terminologia tipica dei metodi bayesiani, e il compimento del passaggio mentale da una soluzione teorica, puramente matematica ed analitica, ad una che sia frutto di simulazioni pratiche ed "empiriche".

Normalmente, riferendosi a modelli a struttura chain-ladder, nessuna ipotesi aprioristica viene effettuata sui parametri di riga. Secondo Bornhuetter-Ferguson, come mostrato all'inizio di questo Capitolo, si ipotizza una certa informazione di base su tali parametri. Tale assunzione è in perfetta analogia con ciò che succede con un approccio Bayesiano. L'informazione nota (o ipotizzata) a priori può essere riassunta con il ricorso ad un'appropriata *prior distribution*, una distribuzione di probabilità aprioristica che descriva i parametri di riga x_i . Sono molte le tipologie di distribuzione a cui far riferimento, anche se probabilmente la miglior scelta ricade sulla distribuzione Gamma:

$$x_i \sim \text{independent Gamma}(\alpha_i, \beta_i)$$

e

$$E(x_i) = U_i^{BF} = \frac{\alpha_i}{\beta_i}$$

Al fine di condurre un paragone tra la tecnica di Bornhuetter-Ferguson e chain-ladder, può essere utile analizzare la distribuzione predittiva (teorica) dei dati. Per i valori incrementali, si parla di una negativa binomiale con sovra-dispersione, con media

$$\left(Z_{ij} D_{i,j-1} + (1 - Z_{ij}) \frac{\alpha_i}{\beta_i \lambda_j \lambda_{j+1} \dots \lambda_n} \right) (\lambda_j - 1)$$

3.6 BOOTSTRAPPING

“Modern computer simulation techniques open up a wide field of practical applications for risk theory concepts, without the restrictive assumptions, and sophisticated mathematics, of many traditional aspects of risk theory”

Daykin, Pentikainen e Pesonen, 1996

Metodi analitici e tecniche di simulazione

Storicamente la variabilità della stima della riserva sinistri è stata determinata facendo ricorso al prediction error, poiché è in effetti piuttosto difficile derivare la distribuzione predittiva completa, ottenuta come somma di variabili casuali. Tuttavia, con l'evoluzione dei calcolatori, computer e software, e lo sviluppo di tecniche di simulazione più robuste, oggi è possibile ottenere una distribuzione simulata di probabilità. Ciò finalmente permette di ovviare all'inconveniente rappresentato dal trattare con formule complicate per il calcolo del prediction error: *“doing it the hard way”* secondo England.

La simulazione, o la generazione di dati artificiali attraverso il computer, può essere usata per molti scopi in statistica; come:

- valutare la variabilità che ci si può attendere in un determinato modello;
- verificare l'adeguatezza di una determinata approssimazione teorica (si pensi all'approssimazione della binomiale con la normale);
- controllare la sensitività dei risultati rispetto alle assunzioni, o per avere indicazioni di ricerca;
- fornire soluzioni numeriche quando non sono disponibili soluzioni analitiche (comprendendo applicazioni in fisica, chimica, ingegneria, finanza, eccetera);

Il primo e più popolare metodo di simulazione è il bootstrapping: un esempio di utilizzo della simulazione per scopi inferenziali, nella fattispecie in ambito frequentista.

Si tratta di una tecnica valida, eppure estremamente semplice; consente di ottenere informazioni da un singolo campione di dati. Di fatto molto simile alla metodologia Monte Carlo, potrebbe addirittura esserne considerata un caso particolare: la differenza risiede nel fatto che, anziché specificare completamente il processo di generazione dei dati (DGP per i modelli Monte Carlo), si usano informazioni dal (l'unico) campione.

Quella appena descritta è l'idea base del bootstrapping; passiamo ora ad osservarne i diversi passaggi nello specifico:

- Il primo step consiste nel ricavare i fattori di sviluppo secondo la tecnica chain-ladder, a partire dai run-off cumulati;
- si ottengono quindi i *fitted values* del triangolo, tramite un procedimento ricorsivo che parte dai valori ultimi osservati. Precisamente

$$\begin{aligned}\widehat{D}_{i,n-i+1} &= D_{i,n-i+1} \\ \widehat{D}_{i,k-1} &= D_{i,k} \frac{1}{\lambda_k};\end{aligned}$$

- si calcolano, per differenza, i *fitted values* incrementali del passato

$$\widehat{m}_{ij} = E(C_{ij});$$

- e successivamente i residui di Pearson, secondo la formula

$$r_{ij}^{(P)} = \frac{C_{ij} - \widehat{m}_{ij}}{\sqrt{\widehat{m}_{ij}}};$$

- si deriva il parametro di scala di Pearson φ , dato dalla somma dei residui al quadrato, divisa per i gradi di libertà; questi sono pari al numero delle osservazioni a cui è stato sottratto il numero di parametri stimati:

$$\varphi = \frac{\sum_{i,j,n-i+1} (r_{ij}^{(P)})^2}{\frac{1}{2}n(n+1) - 2n + 1};$$

- I residui di Pearson vengono dunque aggiustati per replicare la correzione applicata negli approcci analitici

$$r_{ij}^{adj} = r_{ij}^{(P)} \cdot \sqrt{\frac{n}{\frac{1}{2}n(n+1) - 2n + 1}}.$$

- si estraggono campioni dei residui aggiustati. Il campionamento avviene un prefinito numero di volte, con re-immissione dei dati. Si crea dunque un nuovo triangolo dei residui passati;
- per ciascuna delle celle di questo, si esplicitano i valori incrementali C_{ij} , partendo dall'equazione per $r_{ij}^{(P)}$:

$$C_{ij} = r_{ij}^{(P)} \sqrt{\hat{m}_{ij}} + \hat{m}_{ij}.$$

Così facendo si ottiene un set di pseudo-dati in forma incrementale, che sempre descrivono il passato;

- dunque si crea il corrispondente triangolo di pseudo-dati cumulati: in altri termini, al triangolo dei residui ne corrisponde uno di importi cumulati campionati;
- fatto ciò, si procede con il *fitting of the model*, secondo la tecnica scelta, e si completa il triangolo con i valori futuri stimati;
- si ottiene il triangolo relativo al futuro con i valori in forma prima cumulata, poi incrementale per differenza. Questo verrà usato come media nello step successivo;
- per ogni valore, si simula un pagamento a partire dalla distribuzione del processo, con media \tilde{m}_{ij} e varianza $\varphi\tilde{m}_{ij}$;
- si sommano i pagamenti futuri previsti: per anno di origine e complessivi. Il risultato è la stima delle rispettive riserve.

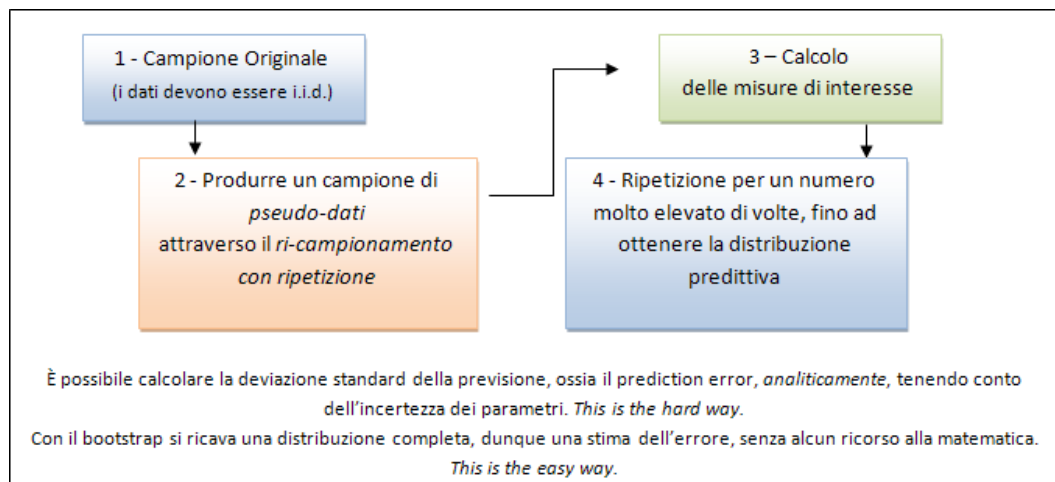


Figura 3.1 – I passaggi del bootstrapping

CONCLUSIONI

Questo secondo Capitolo si contrappone al precedente per la particolare natura dei suoi contenuti: lo sviluppo delle tecnologie ha aperto un sentiero alternativo per la valutazione della riserva sinistri. Se i metodi analitici restano legati alle difficoltà matematiche ed alle complicate formule, la possibilità di replicare la realtà *simulandola* consente sia l'abbandono ed il superamento delle problematiche di calcolo, sia di disporre di preziose informazioni aggiuntive.

Si è innanzitutto inteso introdurre una metodologia differente dal chain-ladder: un utilizzo meccanico di quest'ultimo può, infatti, rischiare di fornire risultati inaffidabili per gli esercizi di sottoscrizione recenti. Ciò, avviene ad esempio quando vengono stimati a zero i sinistri liquidati o denunciati, generando così errori nella stima delle riserve per tali esercizi. L'uso del costo ultimo atteso per gli esercizi più recenti, come avviene con il metodo Bornhuetter-Ferguson, migliora la qualità e la stabilità delle stime.

La statistica bayesiana fa da base e da scuola, coniugando realtà ed aspetti analitici; punto di partenza: le conoscenze a priori, da rivedersi e correggere tramite – e condizionatamente a – l'effettiva esperienza. Da qui, si è passati ai metodi Markov Chain Monte Carlo (MCMC). A tal proposito, alcuni richiami sui processi stocastici con struttura a catena markoviana, e la presentazione dei più comuni algoritmi di calcolo Monte Carlo sono stati effettuati. Queste precisazioni sono in realtà importanti: molti dei software più recenti, come ad esempio WinBUGS (ideato appositamente per l'implementazione di tale tipo di modelli) fanno riferimento ad algoritmi come quello di Gibbs.

Infine, la procedura del bootstrapping, ed in generale i metodi stocastici sono utili per arricchire le informazioni che l'analista ha a disposizione. Tale tecnica rende possibile derivare la distribuzione della variabile aleatoria 'riserva sinistri', adeguandosi in tal maniera ai principi dettati da Solvency II.

L'approccio di stima secondo la chiave dei GLMs fornisce solo medie, varianze e covarianze dei pagamenti futuri per sinistri: si tratta di una limitata serie di valori e momenti. Per ottenere una distribuzione completa degli importi in questione, con la conseguente possibilità di analisi di qualsivoglia misura d'interesse, i modelli possono venir applicati con una procedura di simulazione.

La variabilità della stima è ricavata generando un campione molto ampio di “*pseudo-dati*” – ottenuti appunto per simulazione – partendo dai dati effettivamente osservati. La generazione degli *pseudo-dati* adeguata è ottenuta tramite un algoritmo, sistematizzato da vari lavori scientifici nell’ultimo decennio, e basato sul “rimescolamento” dei residui.

Il primario vantaggio del bootstrap è il superamento che esso consente delle difficoltà analitiche: è pertanto semplicemente implementabile, anche in un normale foglio di calcolo⁴. Ad ogni modo, date le ipotesi di lavoro dell’applicazione standard del bootstrap (vale a dirsi: *l’indipendenza e l’identità di distribuzione* per le variabili da simulare) tale metodo viene sovente applicato ai residui, particolarmente nel caso di modelli di regressione che prevedono l’indipendenza, ma non l’identica distribuzione. È fondamentale quindi la definizione dei residui; ad esempio per il chain-ladder e per il modello ODP si utilizzano i residui di Pearson.

Un’ulteriore fonte di variabilità proviene peraltro dall’uso della stima iniziale dell’indice di sinistralità, richiesta dal metodo. Tutto questo evidenzia l’importanza di acquisire un’approfondita comprensione dei rischi incidenti sui sinistri ben *prima* di applicare qualsiasi metodo in modo meccanico. È solo in seguito che si potrà scegliere e applicare il metodo più appropriato.

Un eccellente testo di approfondimento a cui si rimanda è “*Bootstrap Methodology in Claim Reserving*”, di Pinheiro, Andrade e Centeno, (2001, Astin Colloquium). Infine, se si desidera studiare un modello fuori dal coro, più complesso e pensato apposta per i nuovi principi contabili, si richiama la metodologia di Peter ter Berg, presentata nell’articolo “*Nonlinear Normal Correlated Loss Array: Integrated Financial Modeling of Portfolio and Runoff Risk*” (2001, Astin Colloquium di Washington).

⁴ Nonostante questo, nel Capitolo successivo di questo lavoro l’implementazione del chain-ladder tramite bootstrap è avvenuta con il supporto del software R 2.15.2.

PRATICA - CAPITOLO III

4. I MODELLI E LA LORO APPLICAZIONE ALLA VALUTAZIONE DELLA RISERVA SINISTRI

“You can’t spell Loss Reserving without R”

Brian Fannin

4.1 INTRODUZIONE

Sin dall’inizio del lavoro è stato sottolineato come sia ancora scarso e poco diffuso l’uso di modelli stocastici in determinati settori della statistica attuariale. Ciò, in effetti, lascia sorpresi: da una parte, i recenti sviluppi internazionali chiedono espressamente il ricorso a modelli simili, dall’altra, la quantità di informazioni di cui essi consentono di disporre è nettamente superiore rispetto a quella delle tecniche deterministiche.

In realtà, uno dei maggiori ostacoli alla diffusione dell’utilizzo di questo tipo di metodi si trova nella carenza di software statistici adatti. Pochi sono in commercio, spesso sono difficilmente reperibili e di non semplice utilizzo. Lo testimonia un’indagine condotta dalla *Faculty of Actuaries* nel Regno Unito, riguardante l’impiego dei metodi stocastici nell’attività professionale di claims reserving: tra i principali ostacoli dichiarati figurano la scarsità di risorse, la mancanza di dati e di software adeguati. Tuttavia l’80% degli intervistati avrebbe desiderato utilizzare le metodologie stocastiche più frequentemente, o almeno in talune occasioni.

Nel presente Capitolo si è cercato di offrire un esempio di applicazione pratica dei modelli fin’ora descritti, riferendosi allo stesso set di dati proposto da England e Verrall, particolare poiché caratterizzato dalla presenza di un valore incrementale negativo. Il software su cui è ricaduta la scelta per il supporto dell’analisi è R, noto come la versione ridotta di S-PLUS. Trattasi di un programma estremamente semplice da reperire, ed in continuo costante aggiornamento, poiché *open source*: sfruttare in modo combinato le potenzialità di R e la sua grande diffusione potrebbe rappresentare in futuro un grande stimolo alla diffusione dei modelli stocastici. Lo prova il fatto che, sempre più negli ultimi periodi – ed in particolar modo dalla fine del 2010, con l’entrata in vigore del progetto Solvency II – molti tra gli sviluppatori informatici più importanti si stiano occupando di trovare una soluzione al problema della stima della riserva sinistri.

Per ciascun modello considerato si riassume il suo funzionamento analitico, e le assunzioni teoriche viste in precedenza; vengono in particolar modo evidenziati i passaggi ed i calcoli compiuti grazie alla piattaforma informatica, il loro scopo, i risultati ottenuti. I calcolatori, aggiornati costantemente per essere più veloci e capaci, rendono possibili studi impensabili solo fino a qualche anno fa. Per una maggiore chiarezza espositiva, durante lo svolgimento del Capitolo solo i codici ed i risultati principali sono stati riportati, mentre in appendice al Capitolo è possibile ritrovare lo script completo dei comandi impostati in R.

Da qui in avanti, pertanto, si intende descrivere un metodo di studio e di analisi che potrebbe essere la base di un cambiamento non troppo distante, poiché capace di combinare potenza e semplicità.

4.2 LOSS RESERVING WITH SOFTWARES

4.2.a R Software

R è un software open source particolarmente utile a scopi di analisi, manipolazione e calcolo di dati statistici. Il linguaggio di programmazione è comunemente ritenuto ben sviluppato, semplice ed efficace; la rappresentazione grafica di qualsiasi dato o set di dati è sempre possibile, caratterizzata da facilità e chiarezza. Il software è inoltre piuttosto potente e relativamente veloce.

Più precisamente, si tratta di un ambiente in cui le tecniche statistiche, dalle classiche alle più moderne, sono state implementate e continuano ad esserlo giorno per giorno, grazie alla filosofia sulla condivisione degli sviluppatori. In effetti, pur se numerose funzioni 'di base' sono parte integrante di R, altre sono disponibili in numero sempre crescente grazie ai *packages* disponibili online: continuamente gli esistenti vengono aggiornati, e di nuovi sviluppati. Una volta caricato in R un pacchetto, tutti i suoi contenuti sono disponibili e pronti per essere usati.

La versione del software a cui si fa riferimento è R i386 2.15.2; inoltre, fondamentale per gli scopi del lavoro è stato il pacchetto statistico ChainLadder, nella versione rilasciata a fine 2012.

Il pacchetto ChainLadder

Nel Novembre 2012 è stata rilasciata, da parte degli autori M. Gesmann, W. Zhang e D. Murphy, quella che è attualmente la più recente versione del pacchetto ChainLadder (ChainLadder 0.1.5-4). Il pacchetto è un'estensione pensata per offrire agli utilizzatori diversi dei metodi tipicamente usati per la stima delle riserve sinistri nel ramo danni. L'idea è frutto delle presentazioni tenute da Markus Gesmann presso l'Institute of Actuaries, nel 2007 and 2008, e dai convegni di Dan Murphy e Wayne Zhang per il Casualty Actuarial Society (CAS), rispettivamente nel 2008 e 2010.

Il ChainLadder package può essere scaricato gratuitamente dal sito del progetto CRAN, installato e richiamato in R semplicemente digitando i comandi

```
install.packages('ChainLadder')  
library(ChainLadder)
```

L'ultima versione include attualmente i metodi chain-ladder per i metodi di Mack, di Munich e bootstrap; è possibile trasformare triangoli di run-off cumulati in incrementali e viceversa, triangoli in tavole di dati, eccetera. Una lista di dimostrazioni ed esempi sono compresi nel modello.

Infine, l'ultima interessante funzionalità è la possibilità di utilizzare le funzioni di R in Excel, attraverso l'Add-in scaricabile RExcel.

4.2.b Implementazione dei Modelli

1. I Dati

Nella maggior parte dei casi che ci si propone di analizzare, l'uso di valori cumulati o incrementali risulta essere indifferente. I dati di riferimento sono gli stessi presi in considerazione da Mack nel 1994, e da England e Verrall nell'articolo del 2002: *Automatic Facultative General Liability data (excluding Asbestos and Environmental)*⁵. Le caratteristiche proprie del set ne fanno un caso molto interessante: i pagamenti, presentati in forma incrementale, mostrano una considerevole variabilità tra gli anni di accadimento; si noti in particolar modo la presenza di un valore negativo in una cella del triangolo.

Un secondo motivo a renderlo punto di riferimento è dato dal fatto che lo stesso sia già implementato in R. Nel pacchetto ChainLadder, il set di dati – in forma cumulata – è richiamabile semplicemente con la funzione RAA. Per la trasformazione nella forma incrementale è possibile ricorrere alla funzione `cum2incr(Triangle)`, mentre per ridefinire la forma cumulata si userà `incr2cum(IncTriangle, na.rm=FALSE)`.

Come al solito, la tabella non è che una matrice. Gli elementi nel triangolo superiore sono i valori storici noti, relativi agli anni di origine dal 1981 al 1990, quelli del triangolo inferiore sono invece l'oggetto della stima.

```
data(RAA)
IRAA <- cum2incr(RAA)
RAA
IRAA
```

⁵ *Historical loss development study, 1991.*

Dev Origin	1	2	3	4	5	6	7	8	9	10
1981	5012	8269	10907	11805	13539	16181	18009	18608	18662	18834
1982	106	4285	5396	10666	13782	15599	15496	16169	16704	NA
1983	3410	8992	13873	16141	18735	22214	22863	23466	NA	NA
1984	5655	11555	15766	21266	23425	26083	27067	NA	NA	NA
1985	1092	9565	15836	22169	25955	26180	NA	NA	NA	NA
1986	1513	6445	11702	12935	15852	NA	NA	NA	NA	NA
1987	557	4020	10946	12314	NA	NA	NA	NA	NA	NA
1988	1351	6947	13112	NA	NA	NA	NA	NA	NA	NA
1989	3133	5395	NA	NA	NA	NA	NA	NA	NA	NA
1990	2063	NA	NA	NA	NA	NA	NA	NA	NA	NA

Tabella 1.a – Automatic Facultative General Liability data (1991) – *per valori cumulati*

Dev Origin	1	2	3	4	5	6	7	8	9	10
1981	5012	3257	2638	898	1734	2642	1828	599	54	172
1982	106	4179	1111	5270	3116	1817	-103	673	535	NA
1983	3410	5582	4881	2268	2594	3479	649	603	NA	NA
1984	5655	5900	4211	5500	2159	2658	984	NA	NA	NA
1985	1092	8473	6271	6333	3786	225	NA	NA	NA	NA
1986	1513	4932	5257	1233	2917	NA	NA	NA	NA	NA
1987	557	3463	6926	1368	NA	NA	NA	NA	NA	NA
1988	1351	5596	6165	NA	NA	NA	NA	NA	NA	NA
1989	3133	2262	NA	NA	NA	NA	NA	NA	NA	NA
1990	2063	NA	NA	NA	NA	NA	NA	NA	NA	NA

Tabella 1.b – Automatic Facultative General Liability data (1991) – *per valori incrementali*

I dati possono essere anche letti da R a partire da file in formato .txt, .xls, da internet, e così via; possono inoltre essere inseriti direttamente, ad esempio

```
tri <- t(matrix(
  c(5012, 3257, 2638, 898, 1734, 2542, 1828, 599, 54, 172,
    106, 4179, 1111, 5270, 3116, 1817, -103, 673, 535, NA,
    3410, 5582, 4881, 2268, 2594, 3479, 649, 603, NA, NA,
    5655, 5900, 4211, 5500, 2159, 2658, 984, NA, NA, NA,
```

```

1092, 8473, 6271, 6333, 3786, 225, NA, NA, NA, NA,
1513, 4932, 5257, 1233, 2917, NA, NA, NA, NA, NA,
557, 3463, 6926, 1368, NA, NA, NA, NA, NA, NA,
1351, 5596, 6165, NA, NA, NA, NA, NA, NA, NA,
3133, 2262, NA, NA, NA, NA, NA, NA, NA, NA,
2063, NA, NA, NA, NA, NA, NA, NA, NA, NA),
nc=10, dimnames=list(origin=0:9, dev=0:9))

```

Con la funzione `plot`, si è inoltre in grado di visualizzare graficamente il set di dati;

```
plot(RAA)
```

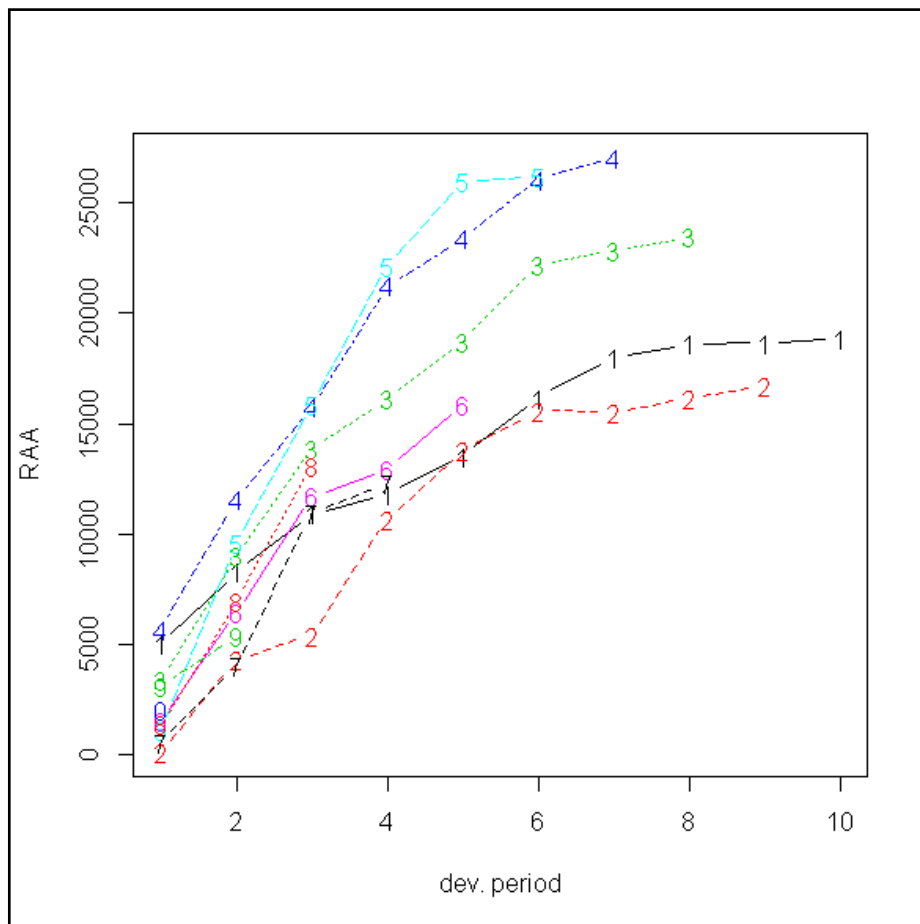


Figura 3.1 – Automatic Facultative General Liability data (1991) – *visualizzazione grafica*

Infine, anche altri triangoli predefiniti sono presenti nel pacchetto, tra questi si ricordano: i run-off di un portafoglio di compensazione di un lavoratore di una grande

compagnia (data(ABC)); un triangolo di run-off incrementali simulati (data(M3IR5)); tre triangoli relativi al *Personal Auto and Commercial Auto Insurance* (data(auto)); i dati di Taylor ed Ashe *General Insurance Claims Data* (data(GenIns)), ecc.

2. Il Chain-Ladder deterministico

Il primo meccanismo da investigare è il classico metodo della catena, la tecnica più semplice, priva di ipotesi sulla distribuzione di partenza, anche se spesso insufficiente ed instabile.

Richiamiamo l'ipotesi fondamentale: la progressione dei pagamenti si mantiene nel tempo, ovvero per ogni generazione; pertanto i link ratio, o rapporti di due colonne adiacenti $w_{ij} = D_{ij}/D_{i,j-1}$ non dipendano da i ma solo da j . I fattori di sviluppo λ *stimano* il rapporto tra le somme di due colonne adiacenti

$$\lambda_j = \frac{\sum_{i=1}^{n-j+1} D_{ij}}{\sum_{i=1}^{n-j+1} D_{i,j-1}}$$

Il metodo chain-ladder è già previsto ed automatizzato nell'omonimo pacchetto di R. Dopo aver richiamato i dati in forma cumulata, è possibile utilizzare due differenti funzioni per ottenere i fattori di sviluppo:

```
chainladder(RAA, weights = 1, delta = 1)
CL <- ChainLadder(RAA)
sapply(CL, coef)
```

λ_j - Development Factors								
2.999	1.6235	1.2708	1.1716	1.1133	1.0419	1.0332	1.0169	1.009

Tabella 2 – Fattori di sviluppo secondo il chain-ladder

Si ottengono in tal modo i valori di $\lambda_2, \lambda_3, \dots, \lambda_n$.

Disponendo di questi, e ricordando che

3. GLMs: Il modello Over-Dispersed Poisson

Si considerino adesso i modelli lineari generalizzati (GLMs). Nel tentativo di restituire le medesime stime del chain-ladder, e di ottenere informazioni aggiuntive oltre la semplice *best estimate*, si implementa il modello di Poisson con sovra-dispersione, che si è già visto essere più appropriato rispetto la sua versione tradizionale.

England e Verrall applicano il modello ODP ai sinistri incrementali C_{ij} . Una volta richiamati in R, è necessaria una preliminare organizzazione dei dati in vettori

```
n.origin <- nrow(IRAA)
n.dvlpmt <- ncol(IRAA)
AY <- factor(row <- rep(1:n.origin, n.dvlpmt))
DY <- factor(col <- rep(1:n.dvlpmt, each=n.origin))
```

In questo modo ogni cella è associata ad un solo AY (accadiment year) ed un solo DY (development year); sono stati creati dei vettori di fattori di riga e colonna in modo tale che per ciascun elemento del vettore CLAIMS, che andiamo a generare, corrisponda un solo elemento dei vettori AY e DY. Inoltre, è utile alla comprensione (anche se non necessario per il calcolo effettivo) visualizzare assieme i tre vettori in un data frame.

```
CLAIMS <- as.vector(IRAA)
mack <- data.frame(claims=CLAIMS, origin=AY, dev=DY)
mack[1:10, ]
```

è bene a questo punto effettuare una considerazione tecnica; quasi tutto è stato predisposto per il fitting del modello, è tuttavia necessario un ultimo aggiustamento. Di norma, è possibile implementare la distribuzione di Poisson con sovra-dispersione tramite la famiglia di funzioni "quasipoisson"; nel set di dati considerato è però presente una cella con valore negativo. Come spiegato nel primo Capitolo della sezione teorica, si tratta di un problema comunque aggirabile, a condizione che la media del modello dopo il fitting sia positiva. Nonostante ciò, R è piuttosto restrittivo a riguardo, e non consente di usare la funzione `quasipoisson`.

Si deve ricorrere ad una funzione alternativa. In questo caso, la soluzione (particolare, e differente di volta in volta) si trova nell'archivio R-help, grazie al contributo di David Firth. Firth ha derivato il seguente codice da sostituire alla funzione:

```
# New quasi-poisson family
quasipoisson <- function (link = "log")
## Amended by David Firth, 2003.01.16, at points labelled ###
## to cope with negative y values
##
## Computes Pearson X^2 rather than Poisson deviance
##
## Starting values are all equal to the global mean
{
  linktemp <- substitute(link)
  if (!is.character(linktemp)) {
    linktemp <- deparse(linktemp)
    if (linktemp == "link")
      linktemp <- eval(link)
  }
  if (any(linktemp == c("log", "identity", "sqrt")))
    stats <- make.link(linktemp)
  else stop(paste(linktemp, "link not available for
poisson",
                 "family; available links are", "\"identity\"", "\"log\"",
and "\"sqrt\""))
  variance <- function(mu) mu
  validmu <- function(mu) all(mu > 0)
  dev.resids <- function(y, mu, wt) wt*(y-mu)^2/mu ###
  aic <- function(y, n, mu, wt, dev) NA
  initialize <- expression({
    n <- rep(1, nobs)
    mustart <- rep(mean(y), length(y)) ###
  })
  structure(list(family = "quasipoisson", link = linktemp,
                linkfun = stats$linkfun, linkinv = stats$linkinv,
                variance = variance,
```

```

      dev.resids = dev.resids, aic = aic, mu.eta =
stats$mu.eta,
      initialize = initialize, validmu = validmu, valideta =
stats$valideta),
      class = "family")
}

```

Effettuata la correzione di Firth, ogni ostacolo al fitting del modello è stato rimosso, consentendo di procedere; i risultati nel dettaglio possono essere osservati tramite la funzione `summary`.

```

model <- glm(claims ~ origin + dev, family = quasipoisson(),
subset=!is.na(claims), data=mack)
summary(model)

```

Call:

```

glm(formula = claims ~ origin + dev, family = quasipoisson(),
data = mack, subset = !is.na(claims))

```

Deviance Residuals:

```

      Min       1Q   Median       3Q      Max
-47.28  -21.20    0.00   15.65   63.13

```

Coefficients:				
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	7.65510	0.30921	24.757	<2e-16 ***
origin2	-0.11084	0.3340	-0.332	0.7420
origin3	0.24586	0.30838	0.797	0.4305
origin4	0.42134	0.30014	1.404	0.1689
origin5	0.42910	0.30300	1.416	0.1653
origin6	0.03481	0.34254	0.102	0.9196
origin7	-0.05932	0.36970	-0.160	0.8734
origin8	0.24319	0.36647	0.664	0.5112
origin9	-0.16027	0.49789	-0.322	0.7494
origin10	-0.02318	0.75658	-0.031	0.9757

dev2	0.69283	0.26000	2.665	0.0115 *
dev3	0.62603	0.26957	2.3220	0.0260 *
dev4	0.27695	0.30164	0.918	0.3647
dev5	0.06056	0.33089	0.183	0.8558
dev6	-0.19582	0.37621	-0.521	0.6059
dev7	-1.08309	0.58871	-1.840	0.0741
dev8	-1.27366	0.76415	-1.667	0.1042
dev9	-1.91593	1.31937	-1.452	0.1551
dev10	-2.50760	2.41131	-1.040	0.3053

Tabella 4 – MODELLO ODP: Stima dei Parametri e loro Standard Error

(Dispersion parameter for quasipoisson family taken to be 983.635)

I risultati sono (quasi) gli stessi ottenuti da England e Verrall. I valori della tabella 4 indicano c, α_i, β_j e φ , ed i rispettivi standard errors. Una volta ottenuti, è semplice sommarli opportunamente (Excel sarà una piattaforma di lavoro sufficiente) per ottenere i fitted values del triangolo dei run-off.

$$\eta_{ij} = c + \alpha_i + \beta_j = \log(m_{ij})$$

$$\hat{C}_{ij} = \hat{m}_{ij} = e^{\hat{\eta}_{ij}}$$

Dev Origin	2	3	4	5	6	7	8	9	10
1982									154
1983								397	220
1984							900	474	262
1985						1098	907	477	264
1986					1797	740	612	322	178
1987				2114	1636	674	557	293	162
1988			3552	2861	2214	912	753	396	219
1989		3364	2373	1911	1479	609	503	265	147
1990	4124	3858	2721	2192	1696	698	577	304	168

Tabella 5 – MODELLO ODP: Fitted Values

La somma di questi valori restituisce la riserva complessiva

fut.design[p, 1 + (n.origin-1) +

```
as.numeric(DY[match(p, fut.points)]) - 1] <- 1
}
```

```
#Design-Matrix:
```

```
fitted.values <- diag(as.vector(exp(fut.design %*% coef)))
total.reserve <- sum(fitted.values)
```

La design matrix dei pagamenti futuri è una matrice quadrata diagonale (in questo caso con 45 righe e 45 colonne), e la riserva totale altro non è che la somma di tutti i fitted values, ossia i valori sulla diagonale. È ora possibile calcolare le covarianze tra i predittori lineari, le covarianze tra i fitted values, e dunque infine il MSEP complessivo, pari a

$$MSEP[\hat{C}_{++}] = \sum \varphi \hat{m}_{ij} + \sum (\hat{m}_{ij})^2 Var(\hat{\eta}_{ij}) + 2 \sum_{fh \neq gk} Cov(\hat{\eta}_{fh}, \hat{\eta}_{gk}) \hat{m}_{fh} \hat{m}_{gk}$$

```
# Covariance matrix of linear predictors
```

```
cov.pred <- fut.design %*% cov.param %*% t(fut.design)
```

```
# Covariance matrix of fitted values
```

```
cov.fitted <- fitted.values %*% cov.pred %*% fitted.values
```

```
#Mean Squared Error of Prediction
```

```
total.rmse <- sqrt(dispen*total.reserve+sum(cov.fitted))
```

```
total.predictionerror <- round(100*total.rmse/total.reserve)
```

Dunque, i risultati del modello possono essere richiamati, riepilogando:

```
total.reserve
```

```
[1] 52135.23
```

```
total.rmse
```

```
[1] 17612.73
```

```
total.predictionerror
```

```
[1] 34
```


7. Bootstrap

Si è più volte ribadito come la procedura del bootstrapping arricchisca di un *quid* importante le informazioni a disposizione dell'analista. Semplifica la valutazione della riserva sinistri, consente di tralasciare i dettagli matematici complessi, mantiene una grande flessibilità e potenza. La novità del bootstrap sta nella possibilità di derivare una completa distribuzione predittiva. *In an easy way.*

In R, i passaggi fondamentali vengono ripercorsi grazie alla funzione `BootChainLadder`, che si occupa in automatico di

- definizione e fitting del modello statistico;
- ottenimento degli pseudo-dati e dei residui;
- riadattamento del modello in base agli pseudo-dati;
- iterazione del loop un elevato numero di volte;
- derivazione delle previsioni e di ogni misura d'interesse.

La funzione, ideata sul modello di England e Verrall in "Stochastic Claims Reserving in General Insurance" (BAJ 2002, Sezione 8) ricorre ad un approccio a due stadi: bootstrapping/simulazione.

In un primo momento l'ordinario metodo del chain-ladder viene applicato ai triangoli dei pagamenti cumulati; a partire da ciò, si calcolano i residui di Pearson aggiustati e su questi agisce il bootstrap, R volte. In questo modo, si ricavano i futuri pagamenti incrementali, in modo analogo al metodo standard.

Nel secondo stadio, la funzione simula l'errore di processo, considerando come media il valore individuato dal bootstrap, e ricorrendo alla distribuzione ipotizzata.

Il set di riserve in tal modo ottenute costituisce la distribuzione predittiva, riassumendo in sé diverse misure statistiche, come ad esempio: media, errore standard, quantili, ecc.

Il comando è strutturato: `BootChainLadder(Triangle, R =999, process.distr =c("gamma", "od.pois"))`

- `Triangle` indica il dataset di partenza;
- `R` è il numero di simulazioni richieste;
- `process.distr`: la distribuzione del processo ipotizzata (gamma o ODP).

Di seguito, l'applicazione del modello al set di dati RAA in R.

Si sono considerate entrambe le distribuzioni Gamma, i cui risultati sono qui riportati, e di Poisson sovra-dispersa, in appendice al Capitolo.

Il numero di simulazioni richieste è stato di un migliaio, soglia comunemente ritenuta affidabile, e particolarmente utilizzata in pratica. Imponendo R =1.000 ed R=10.000, è stato possibile ottenere la distribuzione di probabilità per il Bootstrapping. Invece, con R =100.000, poiché era notevole la mole di lavoro, il software non è riuscito ad elaborare i dati, restituendo un messaggio d'errore.

#Bootstrapping con distribuzione GAMMA

```
Bmodel0<- BootChainLadder(Triangle = RAA, R = 10000, process.distr = "gamma")
```

Errore: cannot allocate vector of size 38.1 Mb

```
Bmodel1<- BootChainLadder(Triangle = RAA, R = 1000, process.distr = "gamma")
```

Bmodel1

	Latest	Mean Ultimate	Mean IBNR	SD IBNR	IBNR 75%	IBNR 95%
1981	18,834	18,834	0	0	0	0
1982	16,704	16,906	202	956	211	1,945
1983	23,466	24,174	708	1,526	1,070	3,577
1984	27,067	28,773	1,706	2,197	2,624	5,868
1985	26,180	29,027	2,847	2,670	4,159	7,889
1986	15,852	19,452	3,600	2,673	4,935	8,486
1987	12,314	18,002	5,688	3,661	7,719	12,380
1988	13,112	24,019	10,907	5,449	14,061	20,743
1989	5,395	16,320	10,925	6,820	14,746	23,294
1990	2,063	19.598	17,535	14,821	25,556	44,317

	Totals
Latest:	160,987
Mean Ultimate:	215,104
Mean IBNR:	54,117
SD IBNR:	20,634
Total IBNR 75%:	65,057
Total IBNR 95%:	92,599

Tabella 18 – BOOTSTRAP con distribuzione Gamma

Mean IBNR indica la stima della riserva sinistri secondo il bootstrapping, SD IBNR è invece il bootstrap prediction error.

Plot(Bmodel1)

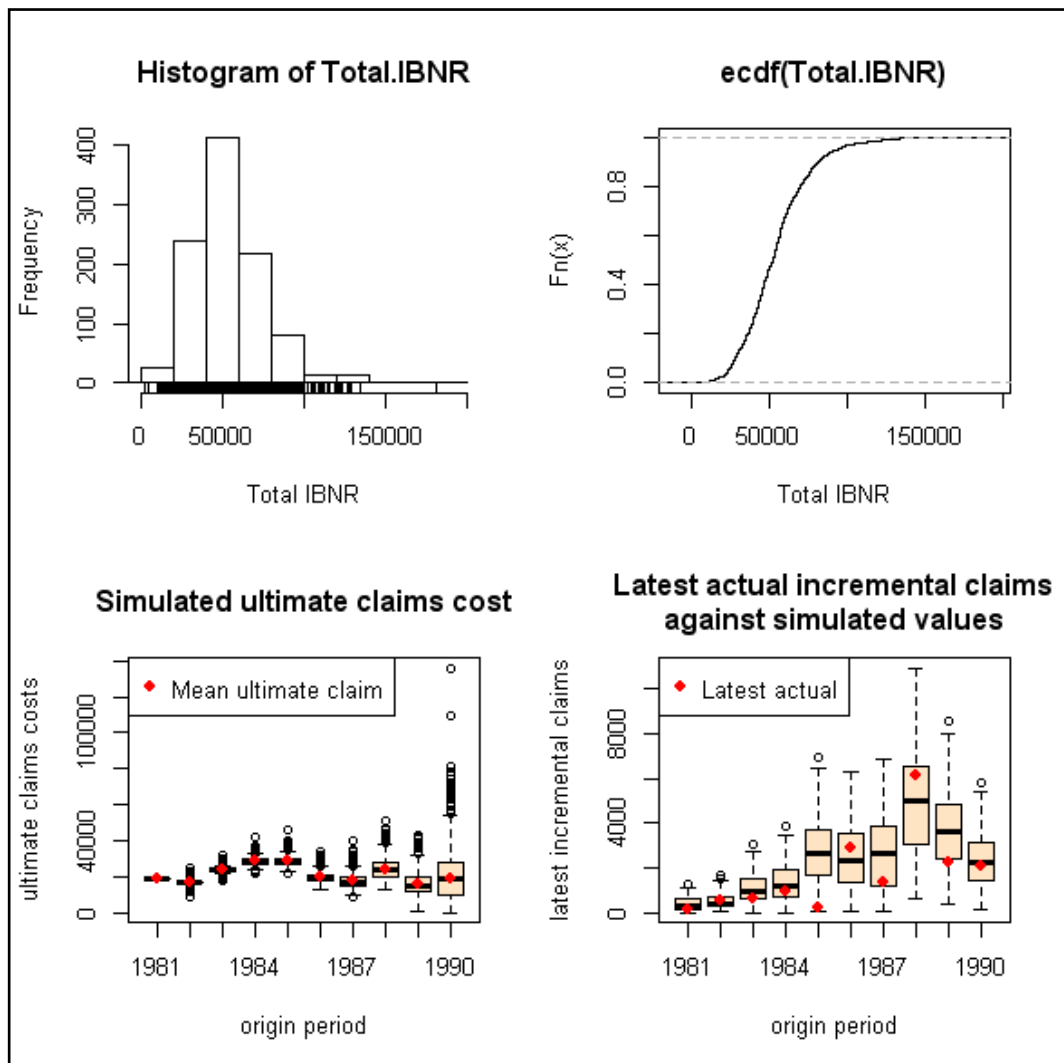


Figura 3.6 – Bootstrapping

I grafici della figura 3.6 indicano rispettivamente (1) l'istogramma delle riserve totali *simulate*, (2) la distribuzione empirica dei totali IBNR, (3) un grafico "box-whisker" dei costi ultimi simulati per anno di origine, (4) un test sull'effettiva possibilità che l'ultimo attuale pagamento incrementale derivi dalla distribuzione simulata dei costi sinistri.

Inoltre, i quantili delle riserve secondo bootstrapping possono essere ricavati con l'apposita funzione `quantile`:

```
quantile(Bmodel1, c(0.75,0.95,0.99, 0.995))
```

\$ByOrigin	IBNR 75%	IBNR 95%	IBNR 99%	IBNR 99,5%
1981	0	0	0	0
1982	211	1,945	3,938	4,433
1983	1,070	3,577	5,352	6,086
1984	2,624	5,868	8,166	9,115
1985	4,159	7,889	11,094	12,106
1986	4,935	8,486	13,517	14,315
1987	7,719	12,380	16,365	18,589
1988	14,061	20,743	28,057	30,023
1989	14,746	23,294	30,351	33,364
1990	25,556	44,317	66,763	72,315
Totals	65,057	92,599	121,404	127,258

Tabella 19 – BOOTSTRAP, quantili della distribuzione

Già attraverso il grafico è possibile notare come la distribuzione delle riserve abbia un andamento log-Normale; pertanto:

```
plot(ecdf(Bmodel1$IBNR.Totals))
fit <- fitdistr(Bmodel1$IBNR.Totals[Bmodel1$IBNR.Totals>0],
"lognormal")
```

5. CONCLUSIONI

Il termine 'assicurare' ha il significato di '*acquistare* la certezza di qualcosa': i soggetti offerenti sono le Compagnie di Assicurazione. Certamente, nessuna è in grado di azzerare la probabilità riguardante il verificarsi di eventi dannosi, e la loro reale offerta garantisce solo contro le ripercussioni economiche che ne scaturirebbero; ciononostante il ruolo giocato è importante e delicato al tempo stesso: *chi acquisterebbe certezza da parte di un venditore verso la cui sicurezza si nutrono dubbi?* Un simile interrogativo impone di far luce sull'affidabilità delle Compagnie stesse. Il raggiungimento di solvibilità, stabilità e sopravvivenza dell'impresa si colloca al centro della rosa di obiettivi del management, degli azionisti, degli assicurati e di ogni stakeholder; un eccellente sistema di riferimento si impernia sulla valutazione delle riserve tecniche, riserva sinistri in primis.

La questione è al centro del dibattito attuariale degli ultimi anni, e lo scenario internazionale è stato palcoscenico di numerosi cambiamenti. Il sentiero indicato dai principi di neutralità, correttezza, di sana e prudente gestione e di materialità dell'informazione conduce al progetto Solvency II (ufficialmente Direttiva 2009/138/CE). In virtù delle innovazioni previste nel modello di business delle Compagnie, la direttiva ha già catalizzato l'attenzione di tutti gli operatori del settore assicurativo. Per non limitarsi ad una gestione degli obiettivi in chiave esclusivamente difensiva, ma anche e soprattutto come opportunità da cogliere, occorre comprendere le logiche sottostanti, e le motivazioni intrinseche del progetto.

*

Strutturato su tre pilastri, Solvency II prevede requisiti patrimoniali minimi, che possano essere aderenti ai rischi fronteggiati generalmente dalle Compagnie Assicuratrici; si prevede inoltre che il calcolo dei suddetti requisiti avvenga secondo una formula standard o, alternativamente, tramite modelli interni di valutazione del rischio. In particolar modo, è nel primo pilastro che i requisiti quantitativi sono ricompresi, accompagnati dai principi circa gli investimenti effettuati dall'assicuratore. Dichiaratamente Solvency II intende incentivare ogni operatore del mercato all'adozione di un modello interno, ad esempio il ricorso alla formula standard comporta maggiori requisiti patrimoniali. Una seconda penalizzazione, non meno importante, è intrinseca

La seconda sezione ha visto così il raggiungimento del cuore del lavoro: una proposta di implementazione pratica dei modelli, con concreti riferimenti informatici.

Sorprende in realtà il dato emergente sull'uso dei modelli stocastici e simulativi: gli sviluppi in ambito internazionale farebbero attendere una percentuale elevata ed in continua crescita, eppure la realtà è ancora ben diversa. La principale spiegazione sta nelle difficoltà di reperimento ed uso di opportuni mezzi informatici; ad esempio, il Gruppo Assicurazioni Generali SpA ricorre ad Igloo, noto per essere una potente suite di modellizzazione del rischio finanziario. Si tratta però di programmi dai costi elevatissimi.

Si è quindi cercato d'offrire un'alternativa più semplice, implementando i principali modelli in R (versione *R i386 2.15.2*). Il pregio di questo software, come è stato più volte detto, sta nel fatto ch'esso sia diffuso, gratuito, e che disponga di innumerevoli estensioni in aggiornamento continuo. Ancora: se in possesso di più che buone conoscenze informatiche e di programmazione, le potenzialità di R lo rendono parificabile anche a software più complessi. Può così essere utilizzato per incontrare le richieste mosse da Solvency II, ottenere risultati risparmiando tempo, permettere all'utilizzatore un'intuitiva comprensione dei dati. Con costanti richiami alla sezione teorica, sono stati illustrati i passaggi, i comandi, ed i principali risultati di una valutazione tipo della riserva sinistri. I pacchetti sviluppati appositamente a fini attuariali consentono poi di ricorrere ad un certo numero di funzioni, in continuo divenire. Alcuni tra i più avanzati, tra cui i packages *actuar* e *ChainLadder*, offrono la possibilità di applicare automaticamente (o con alcuni preliminari aggiustamenti dei dati a disposizione) i più comuni metodi stocastici di riservazione.

*

Si conclude con un monito. Dagli studi effettuati, e già nel corso dello svolgimento del lavoro, è emerso chiaramente come non esista un modello unico, valido dinnanzi a qualsiasi situazione o data-set.

Di volta in volta la scelta dovrà essere accompagnata da criterio e giudizio, poiché solo su tali elementi si può fondare un'analisi che sia accurata e prudente. Questa verificherà la presenza di eventuali peculiarità ovvero anomalie nei dati, e sarà il filo rosso conduttore nell'individuazione del metodo più appropriato.

Bibliografia

- BJÖRKWALL, S. (2011). Stochastic claims reserving in non-life insurance. Bootstrap and smoothing models, Department of Mathematics, Stockholm University;
- BRUNERO, L. (2008). *Introduzione alla statistica bayesiana*, Springer.
- CAVASTRACCI, S. e CENTOFANTI, P.(2005). *Metodi stocastici di valutazione delle riserve sinistri - guida alla lettura*, da Attuari Domani, No. 1-II, pp. 9;
- COLIVICCHI, I. (2011). *Il Bilancio Assicurativo nell'ottica di Solvency 2 e dei Principi IAS/IFRS*. Dipartimento di Matematica per le Decisioni, Università degli Studi di Firenze;
- CUCCHIANI, E.T. (2005). Intervento in *Solvibilità II: vigilanza prudenziale, gestione del rischio, competitività globale*, MIB School of Management, Trieste;
- DABONI, L. (1993). *Lezioni di tecnica attuariale delle assicurazioni contro i danni*, Ed. LINT, Trieste, 6, pp. 227-252;
- DE ALBA, E. and ATONDO, G. (2004). *Claims Reserving When There Are Negative Values in the Runoff Triangle: Bayesian analysis using the three-parameter log-normal distribution*, 39th. Actuarial Research Conference, University of Iowa;
- DE FELICE, M. e MORICONI, F.(2006). *Best estimate delle riserve tecniche e risk margin*, ANIA – Roma;
- DE SILVA, N. (2006). *An Introduction to R: Examples for Actuaries*, Version 0.1, CRAN Project website;
- EFRON, B. and TIBSHIRANI, R.J. (1993). *An Introduction to the Bootstrap*, Chapman and Hall, London;
- ENGLAND, P. (2010). *Stochastic Made "Simple" and Bootstrapping: Lessons Learnt in the Last 10 Years*, EMB, General Insurance Reserving Seminar, Institute of Actuaries;
- ENGLAND, P.D., and VERRALL, R.J. (1998). *Standard Errors of Prediction in Claims Reserving: a comparison of Methods Workshop*, General Insurance Convention & ASTIN Colloquium, Glasgow;
- ENGLAND, P.D., and VERRALL, R.J. (1999). *Comments on: "A comparison of stochastic models that reproduce chain ladder reserve estimates"*, by Mack and Venter Insurance: Mathematics and Economics, 2000, 26/ 1, pp. 109-111;
- ENGLAND, P.D., and VERRALL, R.J. (2002). *A flexible framework for stochastic claims reserving*, *Proceedings of the Casualty Actuarial Society (CAS), Volume LXXXVIII, Part 1 No. 168, May 6, 7, 8, 9, 2001*;

ENGLAND, P.D., and VERRALL, R.J. (2002). *Stochastic Claims Reserving in General Insurance*, British Actuarial Journal, 8/3, pp.443-518;

ENGLAND, P.D., and VERRALL, R.J. (2006). *Predictive distributions of outstanding liabilities in general insurance*, Annals of Actuarial Science, Vol. 1 (2), pp 221-270;

FLYNN, M. and FRANCIS, L.A. (2009). *More Flexible GLMs: Zero-Inflated Models and Hybrid Models*, Casualty Actuarial Society E-Forum;

GESMANN, M. (2011). *Claims Reserving in R, The Chain Ladder Package*, One-day workshop for the Actuarial Profession. 24 July, Staple Inn, London;

GESMANN, M., MURPHY, D. and ZHANG, W. (2012). *R Manual for the ChainLadder Package*, CRAN Project website;

LANDO, B. (2007). *Un modello stocastico per il calcolo della riserva sinistri nell'ottica Solvency II: il metodo di Mack*;

LIU, H. and VERRALL, R.J. (2008). *Bootstrap Estimation of the Predictive Distributions of Reserves Using Paid and Incurred Claims*, Proceedings of the Casualty Actuarial Society, Volume 4, Issue 2;

MACK, T. (1993). *Distribution-free calculation of the standard error of chain ladder reserve estimates*. ASTIN Bulletin. 23/2, pp. 213–225;

MACK, T. (2008). *The Prediction Error of Bornhuetter/Ferguson*, ASTIN Bulletin 38, pp.87-103;

MACK, T. and VENTER, G. (1998). *A comparison of stochastic models that reproduce chain ladder reserve estimates*, Munich Re, Königinstr;

MATARAZZO, L., CAVASTRACCI, S., PASQUALINI, S., DE FELICE, M. e MORICONI, F. (2006). *I metodi stocastici per la misura della riserva sinistri: un approccio al comparto auto italiano*, Sezione Studi ISVAP;

MEYERS, G.G. (2007). *Estimating Predictive Distributions for Loss Reserve Models*, Proceedings of the Casualty Actuarial Society, Volume 1, Issue 2;

PARRINI, C. (2006). *Il filtro di Kalman e la valutazione stocastica della riserva sinistri*, Roma;

PINHEIRO, P.J.R., ANDRADE E SILVA, J.M. and CENTENO, M.L.C. (2003). *Bootstrap methodology in claim reserving*, Journal of Risk and Insurance, 70/4, 701–714;

PIZZOCARO, E. (2008). *La valutazione della riserva sinistri: impiego dei modelli lineari generalizzati*, Milano;

RENSHAW, A.E., VERRALL, R.J. (1998), *A stochastic model underlying the chain ladder technique*, British Actuarial Journal, Vol. 4, pp. 903-923;

SARTORI, N. (2008-2009). *Appunti di Statistica Computazionale II: Metodi di simulazione e Bootstrap*, Padova;

SCOLLNIK, D.P.M. (2001). *Actuarial modelling with MCMC and BUGS*, North American Actuarial Journal, 5 (2), 96-125;

SHAPLAND, M.R. and Leong, J. (2010). *Bootstrap Modeling: Beyond the Basics*, Casualty Actuarial Society E-Forum;

Solvency II Working Party, website;

STRAUB, E. (1988), *Non-Life Insurance Mathematics*, Berlin – Heidelberg – New York, Springer.

TAYLOR, G.C. (2000). *Loss Reserving: An Actuarial Perspective*, Kluwer Academic Publishers, London;

TAYLOR, G.C. and ASHE, F.R. (1983), *Second moments of estimates of outstanding claims*, Journal of Econometrics, Vol. 23, pp. 37-61;

TAYLOR, G.C. and MCGUIRE, G. (2004). *Loss Reserving with GLMs: A Case Study*, Meeting of the Casualty Actuarial Society, Colorado Springs, Colorado;

TAYLOR, G.C., MCGUIRE, G. and GREENFIELD, A. *Loss reserving techniques: past, present and future*, Taylor Fry Consulting Actuaries;

TER BERG, P. (2001). *NONLINEAR NORMAL CORRELATED LOSS ARRAY: Integrated Financial Modeling of Portfolio and Runoff Risk*, ASTIN Colloquium, Washington;

VERRALL, R.J. (2007). *Obtaining Predictive Distributions for Reserves Which Incorporate Expert Opinion*, Variance (Casualty Actuarial Society), 1(1), p.53-80;

VERRALL, R.J. (2000). *An investigation into stochastic claims reserving models and the chain-ladder technique*, Insurance: Mathematics and Economics, 26, 91-99;

VERRALL, R.J. (2004). *A Bayesian generalized linear model for the Bornhuetter-Ferguson method of claims reserving*, North American Actuarial Journal, 8, 67-89;

VERRALL, R.J. and ENGLAND, P.D. (2005). *Incorporating expert opinion into a stochastic model for the chain-ladder technique*. Insurance: Mathematics and Economics, 37, 355-370;

WRIGHT, T.S. (1997). *Probability distribution of outstanding liabilities from individual payment data*, Faculty and Institute of Actuaries, Claims Reserving Manual v.2, Section D7.