

UNIVERSITÀ DEGLI STUDI DI TRIESTE

Dipartimento di Scienze Economiche, Aziendali, Matematiche
e Statistiche

Corso di Laurea Magistrale in Scienze Statistiche e Attuariali

Tesi di Laurea Magistrale

**Modelli stocastici per la valutazione delle
riserve sinistri utilizzando pagamenti e
incurred losses**



Relatore:

Prof.ssa Patrizia Gigante

Candidato:

Fabio CONCINA
matricola: EC7100136

ANNO ACCADEMICO 2014-2015

Indice

Introduzione	1
1 La riserva sinistri	5
1.1 Valutazione della riserva sinistri	5
1.2 Incertezza della previsione	7
1.3 Pagamenti cumulati e incurred losses	8
2 Modello di Mack per il metodo della catena	11
2.1 Modello di Mack	11
2.2 Separate Chain Ladder	14
2.3 Il problema dei P/I	17
3 Modello Munich Chain Ladder	21
3.1 Correlazione tra paid e incurred	21
3.2 Idee preliminari per la soluzione del problema (P/I)	24
3.3 Ipotesi del modello	28
3.4 Analisi delle ipotesi	33
3.5 Stimatori per i parametri	36
3.6 Esempificazioni numeriche	41
4 Modello Paid-Incurred Chain	49
4.1 Ipotesi del modello	49
4.2 Proprietà dei pagamenti cumulati e degli incurred losses	50
4.3 Semplice esemplificazione numerica	60
4.4 Stima dei parametri	61
4.5 Esempificazione numerica	64
5 Modello individuale per Paid and Incurred losses	67
5.1 Dati e notazioni	67
5.2 Il modello	69
5.2.1 Sviluppo dei pagamenti	69
5.2.2 Ipotesi probabilistiche	71
5.3 Calcolo delle riserve	74

6	Valutazioni numeriche	81
6.1	R e il pacchetto <code>ChainLadder</code>	81
6.2	Confronto numerico tra i modelli	83
A	Codice R	87
	Ringraziamenti	97
	Bibliografia	99

Introduzione

Un'impresa assicuratrice che operi nei rami danni deve detenere fondi sufficienti a far fronte al pagamento di sinistri già verificatisi, ma non ancora risarciti o non risarciti in modo completo. Tale accantonamento è detto *riserva sinistri*. La riserva sinistri si forma poiché al verificarsi di un sinistro si origina una passività per l'assicuratore, mentre può intercorrere un periodo di tempo, anche lungo, tra l'epoca di accadimento e l'epoca nella quale il sinistro è risarcito. Inoltre, vi possono essere sinistri già accaduti, ma non ancora denunciati alla compagnia assicuratrice.

Il problema della valutazione della riserva sinistri è complesso ed è affrontato attraverso metodi statistico-attuariali dagli attuari che si occupano della riservazione. Tali metodi utilizzano le informazioni disponibili, relative allo sviluppo dei pagamenti effettuati in passato, e consentono di ottenere in primo luogo una stima della riserva. Alla luce dei nuovi requisiti di solvibilità introdotti con Solvency II, è diventato però sempre più importante l'uso di modelli stocastici per la valutazione delle riserve, poiché questi permettono di ricavare, oltre alla stima puntuale della riserva, anche indicazioni sulla qualità e sulla affidabilità della stima.

Le informazioni disponibili utilizzate nell'ambito dei modelli di riservazione sono generalmente di tipo aggregato e vengono organizzate in tabelle di run-off che riportano sulle righe gli anni di origine, tipicamente gli anni di accadimento dei sinistri, e sulle colonne gli anni di sviluppo. Nelle celle della tabella sono riportati i valori di grandezze di interesse per la valutazione della riserva.

Le grandezze aggregate presenti nelle celle della tabella di run-off possono essere di vario tipo. Principalmente si fa uso di *pagamenti incrementali*, ovvero pagamenti per sinistri di un fissato anno di origine, effettuati con un fissato differimento, e *pagamenti cumulati*, ovvero pagamenti per sinistri di un fissato anno di origine effettuati entro un fissato differimento. Spesso gli assicuratori dispongono però di basi di dati più ricche. In particolare, sono usualmente disponibili le stime degli importi totali dei risarcimenti per sinistri di un fissato anno di origine, valutati alla fine di un fissato anno di differimento. Tali grandezze sono chiamate *incurred losses* e sono date dalla somma tra l'importo totalmente pagato e le stime dell'inventario dell'impegno residuo.

In questo lavoro di tesi è affrontato il problema della valutazione della riserva sinistri nel caso in cui si abbiano a disposizione i dati sui pagamenti cumulati e sugli *incurred losses*. Abbiamo quindi selezionato alcuni modelli che sono stati sviluppati, nella letteratura attuariale, con lo scopo di sfruttare la maggiore ricchezza informativa fornita dalla conoscenza congiunta dei pagamenti cumulati e degli *incurred losses*.

I modelli selezionati sono di tipo stocastico, ovvero le grandezze presenti nella tabella di run-off vengono trattate come variabili aleatorie.

In particolare, abbiamo scelto di analizzare il modello Munich Chain Ladder di Quarg, Mack (2004), il modello Paid-incurred Chain di Merz, Wüthrich (2010) e il modello a dati individuali di Pigeon et al. (2014). Più precisamente, il lavoro è organizzato come segue:

Nel Capitolo 1 è presentato il problema della valutazione della riserva sinistri ed è introdotta l'usuale organizzazione dei dati in tabelle di run-off. E' presentata la differenza tra un approccio deterministico e un approccio stocastico per il calcolo della riserva. Si definisce poi una grandezza, il *mean squared error of prediction* (mse), atta a quantificare l'incertezza legata alla previsione. Sono inoltre descritte le variabili di interesse che saranno utilizzate nel seguito della tesi.

Nel Capitolo 2 è presentato il classico modello di Mack (1993) per il metodo della catena. Sono richiamati gli usuali stimatori dei parametri e del mse per la valutazione dell'incertezza legata alla previsione. E' poi evidenziata la criticità dell'applicazione del modello di Mack nel caso si disponga di dati sia sui pagamenti cumulati e sia sugli *incurred losses*. La criticità è mostrata inizialmente attraverso un esempio numerico e in seguito dimostrata analiticamente.

Nel Capitolo 3 è presentato il modello Munich Chain Ladder, proposto da Quarg, Mack (2004), che ha lo scopo di superare la criticità del modello Chain Ladder messa in evidenza nel capitolo precedente. Sono riproposte, su un diverso insieme di dati, le analisi sviluppate dagli Autori per illustrare la presenza di correlazione tra pagamenti cumulati e *incurred losses* e le considerazioni euristiche che hanno portato alla formulazione delle ipotesi alla base del metodo. E' condotta un'analisi delle ipotesi e sono riportati gli stimatori dei parametri del modello. Sono infine presentate due esemplificazioni numeriche.

Nel Capitolo 4 è presentato il modello Paid-incurred Chain, proposto da Merz, Wüthrich (2010). A differenza del modello Munich Chain Ladder, esso fornisce una stima unica del costo ultimo, utilizzando i dati sui pagamenti cumulati e sugli *incurred losses* e tenendo conto di effetti di correlazione tra i due tipi di variabili. Il metodo sfrutta proprietà della distribuzione Normale multidimensionale e fornisce dei pesi di credibilità per ciascuna fonte d'informazione. Per l'applicazione del metodo è presentato un approccio Bayesiano parziale, nel quale alcuni parametri vengono considerati come deterministici. Ciò ha il vantaggio di permettere il calcolo di forme chiuse per il costo ultimo e per la valutazione della qualità della stima. E' infine presentata una esemplificazione numerica, nella quale si ipotizzano distribuzioni a priori non informative per i parametri.

Nel Capitolo 5 è presentato un modello per *paid e incurred losses*, recentemente proposto da Pigeon et al. (2014), che si applica a dati "individuali". Esso può essere visto come un'estensione del modello Paid-incurred Chain. Il modello utilizza informazioni dettagliate relative ai singoli sinistri di un portafoglio, anziché dati aggregati in triangoli di run-off. Sono introdotte le notazioni, il meccanismo di sviluppo dei pagamenti e le ipotesi probabilistiche del modello. E' poi data la definizione della distribuzione Multivariate Skew Normal, che è usata per il modello, tramite la quale si può tenere conto sia di effetti di correlazione tra le variabili che descrivono i pagamenti e gli *incurred losses*, sia dell'asimmetria delle relative distribuzioni. Sono infine riportati gli stimatori per la valutazione delle riserve di

portafoglio. In questo capitolo non è presente una esemplificazione numerica poiché non ci è stato possibile reperire dati con un così elevato grado di dettaglio.

Nel Capitolo 6 è infine riportata un'applicazione numerica, nella quale sono confrontati i metodi Chain Ladder, Munich Chain Ladder e Paid-incurred Chain su un medesimo insieme di dati, utilizzando il software R. Infine, in Appendice A è riportato il codice R usato per il confronto.

Capitolo 1

La riserva sinistri

La riserva sinistri è una delle principali passività per una impresa assicuratrice che operi nei rami danni. L'esigenza di accantonare tale riserva deriva dal fatto che può intercorrere un periodo di tempo, anche piuttosto lungo, tra l'epoca di accadimento di un sinistro e l'epoca nella quale è pagato il risarcimento, e il sinistro viene chiuso. Il ritardo nel pagamento è dovuto al processo di liquidazione del sinistro, che può essere molto complesso in quanto alcuni sinistri sono di non facile valutazione, possono esserci difficoltà nello stabilire le circostanze e le responsabilità, possono insorgere controversie e procedure legali. Inoltre, può accadere che un sinistro sia denunciato con ritardo rispetto all'epoca di accadimento.

La riserva sinistri, ad una data epoca, è dunque una stima dell'importo che l'assicuratore dovrà pagare in futuro per sinistri che si sono già verificati, ma che non sono stati ancora pagati o non sono stati pagati in modo completo.

Gli attuari che si occupano della riservazione utilizzano metodi statistico-attuariali sviluppati per questo problema specifico. Molti di tali metodi sono storicamente nati come algoritmi deterministici, privi perciò di un preciso inquadramento probabilistico. Sulla base dell'informazione disponibile, tali algoritmi permettono di ricavare un valore puntuale, una 'best estimate', della riserva.

Un limite di questo approccio è l'impossibilità di quantificare l'incertezza legata alla stima, e conseguentemente di valutarne la bontà. I nuovi requisiti di solvibilità impongono inoltre il calcolo di un margine per il rischio da sommare alla 'best estimate' della riserva e di valutare capitali di rischio a fronte del rischio di riservazione. Per poter affrontare tali problemi è necessario l'inquadramento dei metodi di calcolo delle riserve in un ambito probabilistico.

Questo ha motivato l'aumento di interesse riscontrato negli ultimi decenni verso i modelli stocastici per la valutazione delle riserve sinistri.

1.1 Valutazione della riserva sinistri

Con riferimento ai sinistri di un portafoglio omogeneo, l'informazione disponibile è generalmente organizzata in una tabella di run-off. Tale tabella riporta sulle righe gli

anni di origine dei sinistri, che indicheremo con i , e sulle colonne gli anni di sviluppo, che indicheremo con j .

Indichiamo con I il numero di anni di origine per i quali si dispone delle osservazioni e li codifichiamo con $0, \dots, I$, dove I indica l'anno alla fine del quale è valutata la riserva. Si assume che lo sviluppo dei pagamenti rispetto all'anno di origine duri $J + 1$ anni, pertanto $j = 0, \dots, J$. Supponiamo, come è usuale, che riesca $I = J$.

Le osservazioni sono contenute nel triangolo superiore della tabella, detto triangolo di run-off (vedi Tabella 1.1, in grigio il triangolo di run-off). Un'osservazione $y_{i,j}$ è il valore di una grandezza di interesse relativa all'anno di origine i e all'anno di differimento j . Il triangolo superiore della tabella è quindi noto, mentre il triangolo inferiore è oggetto di stima.

Tale impostazione sarebbe sufficiente per l'applicazione dei metodi deterministici di valutazione della riserva: applicando un algoritmo ai dati disponibili è possibile completare il triangolo inferiore e con esso procedere al calcolo della riserva.

Tabella 1.1: tabella di run-off.

$i \setminus j$	0	1	$I - 1$	I
0	$y_{0,0}$	$y_{0,1}$	$y_{0,I}$
1	$y_{1,0}$	$y_{1,1}$	$y_{1,I-1}$	
⋮	⋮	⋮					
i	$y_{i,0}$	$y_{i,1}$...	$y_{i,I-i}$			
⋮	⋮	⋮					
$I - 1$	$y_{I-1,0}$	$y_{I-1,1}$					
I	$y_{I,0}$						

L'impostazione stocastica per la valutazione della riserva sinistri richiede invece di introdurre in corrispondenza con gli elementi della tabella di run-off una famiglia di variabili aleatorie $\{Y_{i,j}, i = 0, \dots, I, j = 0, \dots, I\}$. I dati del triangolo di run-off sono visti come valori osservati dei numeri aleatori $Y_{i,j}$, per $i + j \leq I$. L'importo dei sinistri in sospeso (o *outstanding loss liabilities*), per i quali si accantona la riserva sinistri, è quindi una variabile aleatoria.

Ciascun modello stocastico per il calcolo della riserva sinistri formula delle ipotesi probabilistiche per i numeri aleatori $Y_{i,j}$. Generalmente, tali ipotesi specificano la struttura di alcuni momenti delle $Y_{i,j}$ oppure assumono che la distribuzione di probabilità delle $Y_{i,j}$ appartenga a una assegnata famiglia parametrica. In entrambi i casi, prima di procedere al calcolo della riserva, occorre stimare alcuni parametri.

Per la valutazione della riserva abbiamo bisogno di valutare i numeri aleatori del triangolo inferiore e, in generale, di valutarne una funzione, ovvero di valutare una variabile aleatoria $W = f(Y_{i,j}, i + j > I)$. Pur avendo stimato la distribuzione o i momenti delle variabili $Y_{i,j}$, può non essere banale calcolare la distribuzione o i momenti di W . Per questo motivo, si procede usualmente secondo il seguente approccio: si costruisce uno stimatore

\tilde{W} funzione dei numeri aleatori del triangolo superiore, ovvero una variabile aleatoria $\tilde{W} = g(Y_{i,j}, i + j \leq I)$. Essendo le variabili aleatorie del triangolo superiore osservate è possibile ottenere il valore osservato $\hat{W} = g(y_{i,j}, i + j \leq I)$ di tale stimatore; esso fornisce una previsione della variabile aleatoria W .

1.2 Incertezza della previsione

Per valutare l'incertezza della previsione è possibile utilizzare la radice quadrata del Mean Squared Error of Prediction (msep), il quale è definito come

$$\text{msep}(\tilde{W}) = \mathbf{E}\left((W - \tilde{W})^2\right);$$

è inoltre possibile considerare la radice quadrata del msep condizionato, definito da

$$\text{msep}_{\mathcal{D}_I}(\tilde{W}) = \mathbf{E}\left((W - \tilde{W})^2 | \mathcal{D}_I\right),$$

con $\mathcal{D}_I = \{Y_{i,j}, i + j \leq I\}$.

Osservazione. Se W e \tilde{W} sono stocasticamente indipendenti abbiamo per il msep non condizionato

$$\text{msep}(\tilde{W}) = \mathbf{E}\left((W - \tilde{W})^2\right) = \mathbf{Var}(W) + \mathbf{E}\left((\tilde{W} - \mathbf{E}(W))^2\right), \quad (1.1)$$

in cui

- $\mathbf{Var}(W)$ tiene conto della "variabilità" intrinseca della grandezza aleatoria oggetto di previsione;
- $\mathbf{E}\left((\tilde{W} - \mathbf{E}(W))^2\right)$ tiene conto della "variabilità" della previsione di W ottenuta con lo stimatore \tilde{W} .

Se inoltre lo stimatore \tilde{W} è non distorto, la formula (1.1) si riduce a

$$\text{msep}(\tilde{W}) = \mathbf{Var}(W) + \mathbf{Var}(\tilde{W}).$$

In modo analogo è possibile scomporre il msep condizionato; infatti, sotto l'ipotesi che \tilde{W} sia \mathcal{D}_I -misurabile, si ha

$$\text{msep}_{\mathcal{D}_I}(\tilde{W}) = \mathbf{Var}(W | \mathcal{D}_I) + \left(\mathbf{E}(W | \mathcal{D}_I) - \tilde{W}\right)^2.$$

Osserviamo che $\text{msep}_{\mathcal{D}_I}(\tilde{W})$ è una variabile aleatoria, e che

$$\mathbf{E}\left(\text{msep}_{\mathcal{D}_I}(\tilde{W})\right) = \text{msep}(\tilde{W}).$$

Tali indicatori dipendono dai parametri del modello, inoltre il msep condizionato è una variabile aleatoria. Per arrivare ad una valutazione numerica dell'errore di previsione vengono utilizzate approssimazioni e stime dei due msep.

1.3 Pagamenti cumulati e incurred losses

Nel seguito della tesi verranno considerate come variabili di interesse $Y_{i,j}$

- i pagamenti cumulati, che indicheremo con $P_{i,j}$; $P_{i,j}$ è l'importo dei pagamenti per sinistri di origine i effettuati con differimento minore o uguale a j ;
- gli *incurred losses*, che indicheremo con $I_{i,j}$; $I_{i,j}$ è l'importo totale dei risarcimenti per sinistri con anno di origine i , valutato alla fine dell'anno $i + j$. Esso dato dalla somma dell'importo totalmente pagato $P_{i,j}$ e da una stima dell'impegno residuo.

I dati disponibili saranno quindi due triangoli di run-off, uno per i pagamenti cumulati e uno per gli *incurred losses*.

Considerando la tabella di run-off dei pagamenti cumulati (Tabella 1.2), si ha che

- gli elementi appartenenti all'ultima colonna $P_{0,I}, \dots, P_{I,I}$ sono i *costi ultimi* per ciascun anno di origine;
- gli elementi appartenenti alla diagonale del triangolo $P_{0,I}, P_{1,I-1}, \dots, P_{I,0}$ sono il totale dei pagamenti effettuati per ciascun anno di origine fino al tempo I ;
- fissato i , la sottrazione tra l'elemento dell'ultima colonna e l'elemento diagonale fornisce l'importo dei sinistri in sospeso per l'anno i :

$$P_{i,I} - P_{i,I-i} \quad i = 0, \dots, I;$$

- data una stima $\hat{P}_{i,I}$ del costo ultimo per l'anno di origine i , la riserva per tale anno di origine è

$$R_i^P = \hat{P}_{i,I} - P_{i,I-i};$$

- la somma delle riserve di generazione fornisce la riserva totale di portafoglio:

$$R^P = \sum_{i=0}^I R_i^P.$$

Analogamente, considerando la tabella di run-off degli *incurred losses*, si ha

- gli elementi appartenenti all'ultima colonna $I_{0,I}, \dots, I_{I,I}$ sono i costi ultimi per anno di origine;

Tabella 1.2: tabella di run-off dei pagamenti cumulati; in grigio i valori osservati.

$i \backslash j$	0	1	$J-1$	J
0	$P_{0,0}$	$P_{0,J}$
1	$P_{1,0}$	$P_{1,J-1}$	$P_{1,J}$
⋮	⋮						⋮
i	$P_{i,0}$	$P_{i,J-i}$	$P_{i,J}$
⋮	⋮						⋮
$I-1$	$P_{I-1,0}$	$P_{I-1,1}$	$P_{I-1,J}$
I	$P_{I,0}$	$P_{I,J}$

- fissato i , la sottrazione tra l'elemento dell'ultima colonna della tabella degli *incurred losses* e l'elemento diagonale della tabella dei pagamenti fornisce l'importo dei sinistri in sospeso per l'anno i :

$$I_{i,I} - P_{i,I-i} \quad i = 0, \dots, I;$$

- data una stima $\hat{I}_{i,I}$ del costo ultimo per l'anno di origine i , la riserva per tale anno di origine è

$$R_i^I = \hat{I}_{i,I} - P_{i,I-i};$$

- la somma delle riserve di generazione fornisce la riserva totale di portafoglio:

$$R^I = \sum_{i=0}^I R_i^I.$$

Per valutare la riserva sinistri si può dunque stimare il costo ultimo mediante uno stimatore costruito a partire dalla tabella dei pagamenti, $\tilde{P}_{i,I} = g^P(P_{i,j}, i + j \leq I)$, ma anche mediante uno stimatore costruito a partire dalla tabella degli *incurred losses*, $\tilde{I}_{i,I} = g^I(I_{i,j}, i + j \leq I)$.

Nel prossimo capitolo è illustrato il problema che emerge quando le due stime sono ottenute in modo separato, usando il metodo della catena.

Capitolo 2

Modello di Mack per il metodo della catena

In questo capitolo presentiamo il modello di Mack (1993) per il metodo della catena. Si tratta di un modello stocastico, semiparametrico, introdotto per i pagamenti cumulati, quando i dati siano disponibili in forma aggregata in un triangolo di run-off. Il modello fornisce le stesse stime del metodo deterministico della catena o *Chain Ladder* (CL) e dunque è uno dei modelli stocastici che possono essere visti come sottostanti a tale metodo che, come è ben noto, è molto utilizzato nella pratica attuariale. Benché sia stato introdotto per i pagamenti cumulati, il metodo della catena può essere applicato anche ad altre grandezze di interesse collegate con il problema della valutazione della riserva sinistri, purché siano soddisfatte le ipotesi del modello.

Nel caso si disponga delle informazioni sui pagamenti cumulati ed anche sugli *incurred losses*, in forma di triangoli di run-off, si potrebbe pensare di applicare il metodo *Chain Ladder* in modo separato ai due triangoli. In questo capitolo verranno illustrate le debolezze di tale procedura, che è denominata metodo *Separate Chain Ladder* (SCL).

2.1 Modello di Mack

In questo paragrafo richiamiamo brevemente il modello *Chain Ladder* seguendo l'impostazione semi-parametrica presentata in Mack (1993).

Le variabili aleatorie $P_{i,j}$ indicano i pagamenti cumulati relativi all'anno di origine i e anno di sviluppo j , con $i = 0, \dots, I$ e $j = 0, \dots, I$. Sia $\{P_{i,j}, i = 0, \dots, I, j = 0, \dots, I\}$ il processo stocastico dei pagamenti cumulati relativi alla tabella di run-off. Indichiamo con $\mathcal{P}_i(j) = \{P_{i,0}, \dots, P_{i,j}\}$ l'informazione disponibile dell'anno di origine i fino all'anno di sviluppo j . Sia infine $\mathcal{D}_I = \{P_{i,j}, i + j \leq I\}$ l'osservazione disponibile al tempo I .

Ipotesi 2.1 (Modello di Mack).

- (i) I vettori $\{P_{i,1}, \dots, P_{i,I}\}$ sono stocasticamente indipendenti al variare di i ;

(ii) Esistono $f_0, \dots, f_{I-1} > 0$ tali che

$$\mathbf{E}(P_{i,j+1}|\mathcal{P}_i(j)) = \mathbf{E}(P_{i,j+1}|P_{i,j}) = P_{i,j} \cdot f_j, \quad j = 0, \dots, I-1, \quad i = 0, \dots, I;$$

(iii) Esistono $\sigma_0, \dots, \sigma_{I-1} > 0$ tali che

$$\mathbf{Var}(P_{i,j+1}|\mathcal{P}_i(j)) = \mathbf{Var}(P_{i,j+1}|P_{i,j}) = P_{i,j} \cdot \sigma_j^2, \quad j = 0, \dots, I-1, \quad i = 0, \dots, I.$$

□

Il seguente risultato permette di ottenere la speranza matematica del costo ultimo per l'anno di origine i , $P_{i,I}$, condizionata all'informazione sui pagamenti disponibili in I . Tale speranza matematica è lo stimatore bayesiano del costo ultimo.

Lemma 2.1. *Sotto le Ipotesi 2.1 abbiamo*

$$\mathbf{E}(P_{i,I}|\mathcal{D}_I) = \mathbf{E}(P_{i,I}|P_{i,I-i}) = P_{i,I-i} \cdot f_{I-i} \cdots f_{I-1}, \quad i = 1, \dots, I.$$

Dimostrazione. Si ha

$$\begin{aligned} \mathbf{E}(P_{i,I}|\mathcal{D}_I) &= \mathbf{E}(P_{i,I}|P_{i,0}, \dots, P_{i,I-i}) = \\ &= \mathbf{E}\left(\mathbf{E}(P_{i,I}|P_{i,0}, \dots, P_{i,I-i}, P_{i,I-i+1}, \dots, P_{i,I-1})|P_{i,0}, \dots, P_{i,I-i}\right) = \\ &= \mathbf{E}\left(\mathbf{E}(P_{i,I}|P_{i,I-1})|P_{i,0}, \dots, P_{i,I-i}\right) = \\ &= f_{I-1} \cdot \mathbf{E}(P_{i,I-1}|P_{i,0}, \dots, P_{i,I-i}), \end{aligned}$$

dove la prima uguaglianza segue dall'indipendenza dei pagamenti relativi a diversi anni di origine e la seconda dalla proprietà iterativa¹ della speranza matematica. Le ultime due uguaglianze seguono dall'ipotesi (ii). Iterando il ragionamento si ottiene la tesi. □

Uno stimatore della riserva per l'anno di origine i è allora

$$\tilde{R}_i = \mathbf{E}(P_{i,I}|\mathcal{D}_I) - P_{i,I-i} = P_{i,I-i} \cdot (f_{I-i} \cdots f_{I-1} - 1), \quad i = 1, \dots, I.$$

Per $i = 0$ la riserva è nulla, infatti, essendo $P_{0,I}$ \mathcal{D}_I -misurabile si ha $\mathbf{E}(P_{0,I}|\mathcal{D}_I) - P_{0,I} = P_{0,I} - P_{0,I} = 0$.

Per l'applicazione del modello è ora necessario introdurre degli stimatori per i parametri f_j , che sono detti fattori di sviluppo o *link ratios*. Si prova che i seguenti stimatori sono non distorti e non correlati (si veda Mack (1993) e Merz, Wüthrich (2008)):

¹Ricordiamo che se X è un numero aleatorio di speranza matematica finita, allora se \mathcal{H} è una sovra σ -algebra di una σ -algebra \mathcal{K} , riesce $\mathbf{E}(X|\mathcal{K}) = \mathbf{E}[\mathbf{E}(X|\mathcal{H})|\mathcal{K}]$.

$$\tilde{f}_j = \frac{\sum_{i=0}^{I-j-1} P_{i,j+1}}{\sum_{i=0}^{I-j-1} P_{i,j}}, \quad j = 0, \dots, I-1.$$

Lo stimatore del costo ultimo $P_{i,I}$ è allora

$$\tilde{P}_{i,I} = P_{i,I-i} \cdot \tilde{f}_{I-i} \cdots \tilde{f}_{I-1}, \quad i = 1, \dots, I,$$

mentre gli stimatori delle riserve per diversi anni di origine sono

$$\tilde{R}_i = \tilde{P}_{i,I} - P_{i,I-i}, \quad i = 1, \dots, I.$$

Lo stimatore della riserva totale è infine

$$\tilde{R} = \sum_{i=0}^I \tilde{R}_i,$$

con $\tilde{R}_0 = 0$.

Sostituendo nelle espressioni degli stimatori i valori osservati dei pagamenti cumulati del triangolo di run-off si ottengono le stime. Le stime coincidono con quelle del CL deterministico.

Ai fini di valutare l'errore di previsione è necessario stimare anche i parametri di varianza σ_j^2 , $j = 0, \dots, I$. Si prova che i seguenti stimatori sono non distorti:

$$\tilde{\sigma}_j^2 = \frac{1}{I-j-1} \sum_{i=0}^{I-j-1} P_{i,j} \left(\frac{P_{i,j+1}}{P_{i,j}} - \tilde{f}_j \right)^2, \quad j = 0, \dots, I-2.$$

Per stimare σ_{I-1}^2 è usato il seguente stimatore

$$\tilde{\sigma}_{I-1}^2 = \min \left\{ \tilde{\sigma}_{I-2}^4 / \tilde{\sigma}_{I-3}^2; \tilde{\sigma}_{I-3}^2; \tilde{\sigma}_{I-2}^2 \right\}.$$

Nel seguente lemma riportiamo degli stimatori per il msep condizionato dello stimatore del costo ultimo per singolo anno di origine e per anni di origine aggregati. Per le relative dimostrazioni si veda Merz e Wüthrich (2008).

Lemma 2.2. *Sotto le Ipotesi 2.1 uno stimatore per $\text{mse}_{P_{i,I}|\mathcal{D}_I}(\tilde{P}_{i,I}) = \text{mse}_{R_i|\mathcal{D}_I}(\tilde{R}_i)$ è dato da*

$$\begin{aligned} \widetilde{\text{mse}}_{P_{i,I}|\mathcal{D}_I}(\tilde{P}_{i,I}) &= \tilde{P}_{i,I}^2 \cdot \sum_{j=I-i}^{I-1} \frac{\tilde{\sigma}_j^2/\tilde{f}_j^2}{\tilde{P}_{i,j}} \\ &+ P_{i,I-i}^2 \cdot \prod_{j=I-i}^{I-1} \tilde{f}_j^2 \sum_{j=I-i}^{I-1} \frac{\tilde{\sigma}_j^2/\tilde{f}_j^2}{\sum_{i=0}^{I-j-1} P_{i,j}}, \quad i = 1, \dots, I. \end{aligned}$$

Uno stimatore per $\text{mse}_{\sum_i P_{i,I}|\mathcal{D}_I}(\sum_{i=1}^I \tilde{P}_{i,I}) = \text{mse}_{R|\mathcal{D}_I}(\tilde{R})$ è dato da

$$\begin{aligned} \widetilde{\text{mse}}_{\sum_i P_{i,I}|\mathcal{D}_I}\left(\sum_{i=1}^I \tilde{P}_{i,I}\right) &= \sum_{i=1}^I \widetilde{\text{mse}}_{P_{i,I}|\mathcal{D}_I}(\tilde{P}_{i,I}) \\ &+ 2 \sum_{1 \leq i < k \leq I} \tilde{P}_{i,I} \cdot \tilde{P}_{k,I} \cdot \sum_{j=I-i}^{I-1} \frac{\tilde{\sigma}_j^2/\tilde{f}_j^2}{\sum_{i=0}^{I-j-1} P_{i,j}}. \end{aligned}$$

2.2 Separate Chain Ladder

Come si è detto nell'introduzione, il metodo della catena è stato introdotto per i pagamenti cumulati, tuttavia può essere applicato a qualunque grandezza di interesse $Y_{i,j}$ relativa alle celle di una tabella di run-off, purché siano soddisfatte le ipotesi del modello sottostante.

Nella pratica attuariale il metodo è spesso applicato anche agli *incurred losses*. Infatti, se è possibile accogliere per gli *incurred losses* ipotesi analoghe alle Ipotesi 2.1, si può completare la tabella di run-off relativa agli *incurred losses* usando l'algoritmo della catena, ottenendo in particolare le stime dei costi ultimi $\hat{I}_{i,I}$. Accade però che stimando i costi ultimi, e quindi le riserve, usando separatamente i dati sui pagamenti cumulati $P_{i,j}$ e sugli *incurred losses* $I_{i,j}$ si ottengono stime spesso sensibilmente diverse. Tale approccio è detto in Quarg, Mack (2004) *Separate Chain Ladder* (SCL).

Nel seguente esempio ripercorriamo le analisi sviluppate da Quarg, Mack (2004), su una diversa base di dati, per evidenziare le criticità del metodo SCL. Un'idea chiave sarà considerare i rapporti (P/I) tra pagamenti e incurred losses definiti come

$$(P/I)_{i,j} := \frac{P_{i,j}}{I_{i,j}}, \quad i, j = 0, \dots, I.$$

Esempio. Per le valutazioni seguenti si è utilizzato il linguaggio R e il pacchetto `ChainLadder`, come verrà descritto in dettaglio nel Capitolo 6. Consideriamo i dati di un portafoglio americano di assicurazioni di responsabilità civile auto, per autoveicoli adibiti ad uso privato, della compagnia USAA (United Services Automobile Association)². Abbiamo a

²Fonte http://www.casact.org/research/index.cfm?fa=loss_reserves_data.

disposizione i triangoli di run-off dei pagamenti e degli *incurred losses* con $I = 9$. Nella Tabella 2.1 sono riportate le stime dei costi ultimi e delle riserve per anno di origine ottenute con i due triangoli di dati, mentre nella Tabella 2.2 sono riportate le stime delle riserve totali \hat{R}^P e \hat{R}^I . E' evidente che c'è una differenza importante tra le due riserve, mentre si vorrebbero avere risultati per quanto possibile simili. Questo è un primo problema che si riscontra nell'applicazione del metodo SCL.

Tabella 2.1: stime dei costi ultimi e delle riserve per generazione usando SCL.

i	$\hat{P}_{i,I}$	$\hat{I}_{i,I}$	\hat{R}_i^P	\hat{R}_i^I
0	886,334	893,764	0	7,430
1	982,932	985,296	784	3,148
2	1,079,102	1,077,957	3,565	2,420
3	1,146,697	1,135,524	8,322	-2,851
4	1,246,146	1,226,920	19,496	270
5	1,375,073	1,333,337	50,341	8,605
6	1,440,528	1,367,608	120,398	47,478
7	1,430,755	1,289,717	245,455	104,417
8	1,455,908	1,222,449	489,746	256,287
9	1,568,804	1,065,258	1,026,783	523,237

Tabella 2.2: stime delle riserve totali usando SCL.

\hat{R}^P	\hat{R}^I
1,964,890	950,441

Calcoliamo ora il triangolo dei rapporti $(P/I)_{i,j}$ e costruiamo un grafico ponendo sulle ascisse gli anni di sviluppo e sulle ordinate tali rapporti (Figura 2.1). La linea tratteggiata indica la media ponderata dei rapporti per anno di sviluppo, usando come pesi gli *incurred losses* $I_{i,j}$.

Si può notare che i rapporti (P/I) sono dispersi attorno alla media, ma la dispersione decresce al crescere degli anni di sviluppo. Inoltre, vi è la tendenza a raggiungere il valore 1.0 dopo 10 anni di sviluppo.

Proiettiamo ora separatamente per mezzo del metodo SCL il triangolo dei pagamenti cumulati e il triangolo degli *incurred losses*. Calcoliamo i rapporti $(P/I)_{i,j}$ proiettati e

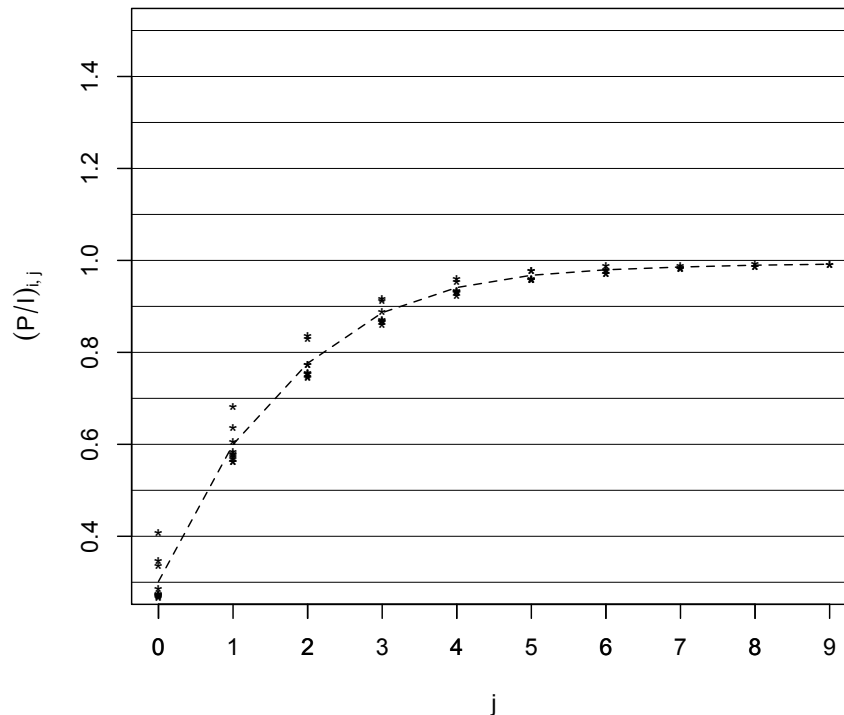


Figura 2.1: rapporti (P/I) e anni di sviluppo.

costruiamo un grafico analogo al precedente (Figura 2.2).

Dall'analisi del grafico risultano evidenti i seguenti fatti:

- la dispersione dei rapporti (P/I) attorno alla media è crescente al crescere degli anni di sviluppo, diversamente da quanto osservato nei dati;
- per l'ultimo anno di sviluppo abbiamo rapporti che vanno da 99.17% a 147.27%, indicando che la proiezione dei pagamenti è circa uguale a quella degli *incurred losses* in alcuni anni di origine, mentre è molto più elevata in altri.

Nel grafico in Figura 2.3 abbiamo indicato in nero i rapporti correnti $(P/I)_{i,I-i}$ e in grigio i rapporti proiettati. Ciascun simbolo corrisponde a un anno di origine (sono stati riportati solo gli anni di origine $i = 3,5,7,8,9$ per evitare sovrapposizioni). Il grafico evidenzia la seguente tendenza: rapporti correnti $(P/I)_{i,I-i}$ superiori alla media sono proiettati in rapporti superiori alla media. Nell'esempio considerato si ha che tutti i rapporti correnti $(P/I)_{i,I-i}$ sono superiori alla media. Nel paragrafo seguente verrà dimostrato che avviene anche il viceversa, ovvero che rapporti correnti inferiori alla media sono proiettati in rapporti inferiori alla media.

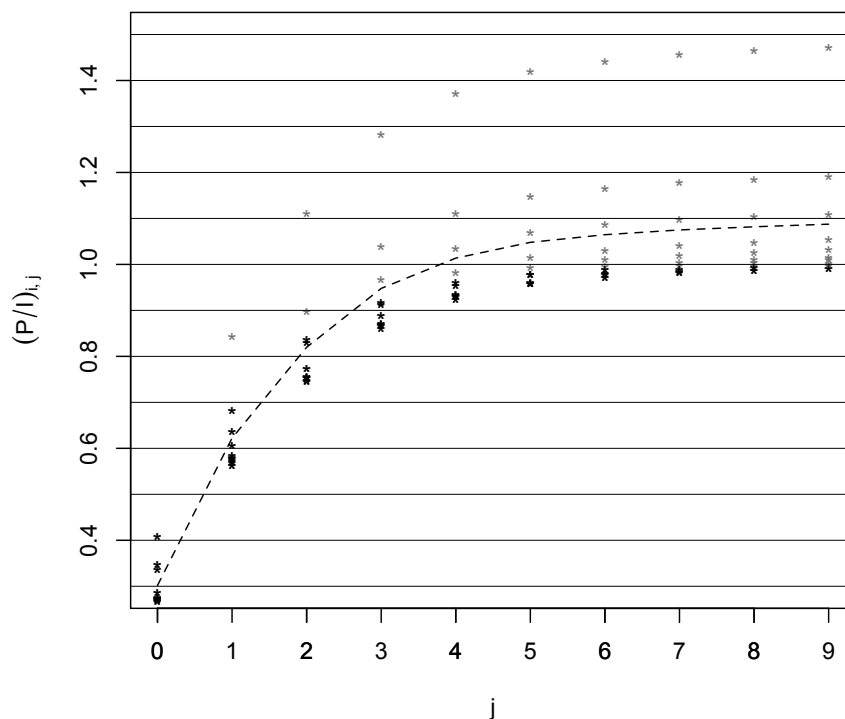


Figura 2.2: rapporti (P/I) e anni di sviluppo. In grigio i rapporti proiettati.

Queste sono debolezze del metodo e hanno come conseguenza una non coerenza delle previsioni con i dati osservati.

□

Nel prossimo paragrafo verrà dimostrato formalmente che il problema dei rapporti (P/I) , evidenziato nell'esempio, è una caratteristica strutturale e ineliminabile del metodo SCL.

2.3 Il problema dei P/I

Per spiegare ciò che accade proiettando separatamente i pagamenti cumulati e gli *incurred losses*, consideriamo il modello deterministico per il metodo della catena. Ciò è giustificato dal fatto che con tale modello e con il modello stocastico si ottengono le medesime proiezioni.

Nel metodo deterministico della catena, per i due tipi di dati, si assume che

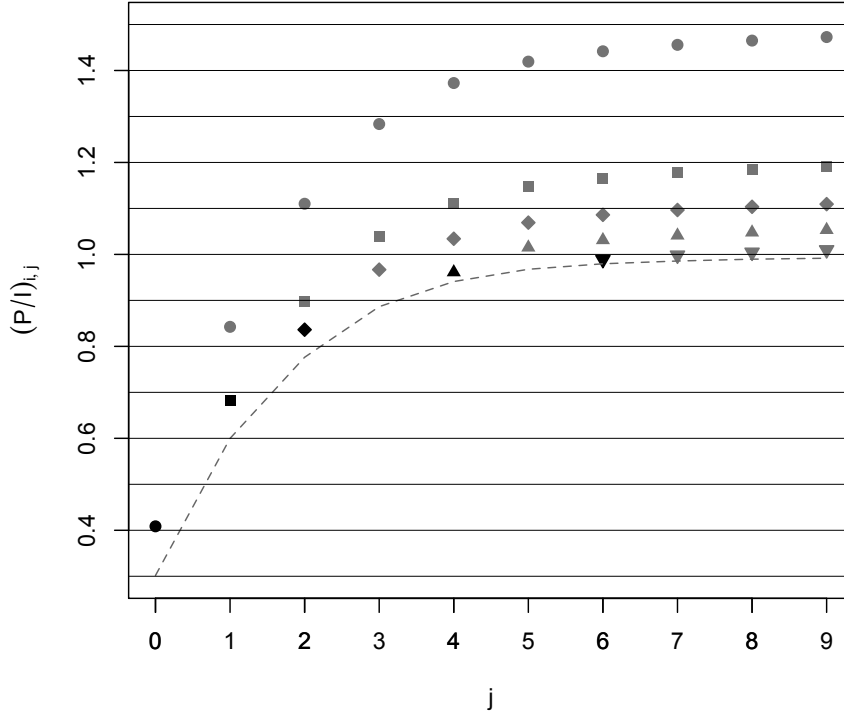


Figura 2.3: rapporti correnti $(P/I)_{i,I-i}$ superiori alla media vengono proiettati in rapporti superiori alla media.

$$\frac{P_{i,j+1}}{P_{i,j}} = f_j^P, \quad \frac{I_{i,j+1}}{I_{i,j}} = f_j^I, \quad i = 0, \dots, I, \quad j = 0, \dots, I-1. \quad (2.1)$$

Si ha dunque

$$P_{i,j} = P_{i,I-i} \cdot f_{I-i}^P \cdots f_{j-1}^P, \quad j > I-i,$$

e

$$I_{i,j} = I_{i,I-i} \cdot f_{I-i}^I \cdots f_{j-1}^I, \quad j > I-i.$$

Osserviamo che, dalla (2.1), si ha

$$\sum_{i \in \mathcal{J}} P_{i,j+1} = f_j^P \cdot \sum_{i \in \mathcal{J}} P_{i,j},$$

per ogni insieme $\mathcal{J} \subset \{0, \dots, I\}$. Dunque si ha

$$f_j^P = \frac{\sum_{i=0}^{I-j-1} P_{i,j+1}}{\sum_{i=0}^{I-j-1} P_{i,j}},$$

ma anche

$$f_j^P = \frac{\sum_{i=0}^I P_{i,j+1}}{\sum_{i=0}^I P_{i,j}}, \quad j = 0, \dots, I-1. \quad (2.2)$$

Analoghe considerazioni sussistono per f_j^I . Pertanto

$$f_j^I = \frac{\sum_{i=0}^I I_{i,j+1}}{\sum_{i=0}^I I_{i,j}}, \quad j = 0, \dots, I-1. \quad (2.3)$$

Consideriamo ora la media ponderata $(P/I)_j$ dei rapporti (P/I) per l'anno di sviluppo j , al variare di i , definita da

$$(P/I)_j := \frac{1}{\sum_{i=0}^I I_{i,j}} \cdot \sum_{i=0}^I I_{i,j} \cdot (P/I)_{i,j} = \frac{\sum_{i=0}^I P_{i,j}}{\sum_{i=0}^I I_{i,j}}, \quad j = 0, \dots, I,$$

e i rapporti proiettati (P/I)

$$(P/I)_{i,j} = \frac{P_{i,I-i} \cdot f_{I-i}^P \cdot \dots \cdot f_{j-1}^P}{I_{i,I-i} \cdot f_{I-i}^I \cdot \dots \cdot f_{j-1}^I}, \quad j > I-i, \quad i = 0, \dots, I. \quad (2.4)$$

Sostituendo nella formula (2.4) le (2.2), (2.3), otteniamo

$$(P/I)_{i,j} = \frac{P_{i,I-i} \cdot \frac{\sum_{h=0}^I P_{h,j}}{\sum_{h=0}^I P_{h,I-i}}}{I_{i,I-i} \cdot \frac{\sum_{h=0}^I I_{h,j}}{\sum_{h=0}^I I_{h,I-i}}},$$

che possiamo riscrivere come

$$(P/I)_{i,j} = \frac{P_{i,I-i} \cdot \sum_{h=0}^I P_{h,j}}{I_{i,I-i} \cdot \sum_{h=0}^I I_{h,j}}.$$

Otteniamo dunque la relazione

$$\frac{(P/I)_{i,j}}{(P/I)_j} = \frac{(P/I)_{i,I-i}}{(P/I)_{I-i}}, \quad j > I - i. \quad (2.5)$$

Il significato della relazione (2.5) è il seguente: per ogni anno di origine, il rapporto tra un valore (P/I) proiettato e la corrispondente media è uguale al rapporto tra il valore (P/I) corrente e la rispettiva media. Un anno di origine con rapporto $(P/I)_{i,I-i}$ corrente sopra la media avrà rapporti proiettati $(P/I)_{i,j}$ sopra la media; viceversa per un rapporto corrente sotto la media. Un tale comportamento delle stime era stato evidenziato nell'esempio sviluppato nel Paragrafo 2.2. Da quanto qui osservato si ha che esso è legato in modo intrinseco al metodo SCL.

Nel capitolo successivo verrà identificata la causa della debolezza di tale metodo e verranno proposte delle idee preliminari per il suo superamento. Tali considerazioni verranno poi formalizzate nelle ipotesi alla base del metodo Munich Chain Ladder (MCL) come proposto in Quarg e Mack (2004), il quale permette di superare i limiti del metodo SCL.

Capitolo 3

Modello Munich Chain Ladder

In questo capitolo verranno definite e analizzate le ipotesi alla base del metodo Munich Chain Ladder (MCL) di Quarg, Mack (2004) e verranno definiti degli stimatori per permettere l'applicazione del metodo proposto. Preliminarmente verrà illustrata la causa del problema evidenziato nel capitolo precedente, relativo all'andamento dei rapporti (P/I) , quando si applichi il metodo SCL. Le analisi sviluppate nei primi due paragrafi ripercorrono le analisi dell'articolo citato, ma utilizzando diversi triangoli di dati.

3.1 Correlazione tra paid e incurred

Alla base della debolezza del metodo SCL vi è ignorare la correlazione tra pagamenti cumulati (*paid*) e *incurred losses*. Per trattare il problema della correlazione, consideriamo i fattori di sviluppo individuali, definiti come rapporti tra pagamenti cumulati, o *incurred losses*, consecutivi, per ciascun anno di origine:

$$f_{i,j}^P = \frac{P_{i,j+1}}{P_{i,j}}; \quad f_{i,j}^I = \frac{I_{i,j+1}}{I_{i,j}}.$$

Si tratta dei rapporti dei quali si considera la media per stimare i fattori di sviluppo f_j^P , f_j^I del metodo della catena.

Il prossimo esempio illustra la presenza di dipendenza tra i fattori di sviluppo individuali e i rapporti (P/I) .

Esempio. In questo esempio considereremo i dati in Wüthrich e Merz (2008). Si hanno a disposizione due triangoli di run-off di dimensione $I = 15$, il primo con pagamenti cumulati e il secondo con *incurred losses*. Calcoliamo i fattori di sviluppo individuali per i pagamenti, relativi al passaggio tra il differimento 0 e il differimento 1, $f_{i,0}^P$, e i rapporti $(P/I)_{i,0}$, ovvero

$$f_{i,0}^P = \frac{P_{i,1}}{P_{i,0}} \quad \text{e} \quad (P/I)_{i,0} = \frac{P_{i,0}}{I_{i,0}}, \quad i = 0, \dots, I - 1.$$

Nella Tabella 3.1 sono riportati i risultati. Oltre ai valori $f_{i,0}^P$, $(P/I)_{i,0}$, sono riportati gli scarti dalle rispettive medie

$$\Delta_1 = f_{i,0}^P - \bar{f}_0^P, \quad \Delta_2 = (P/I)_{i,0} - \overline{(P/I)}_{i,0},$$

Tabella 3.1: confronto tra fattori di sviluppo individuali $f_{i,0}^P$ e rapporti $(P/I)_{i,0}$.

i	$f_{i,0}^P$	$(P/I)_{i,0}$	$\Delta_1\%$	$\Delta_2\%$
0	1.3861	0.6560	-5.52%	3.44%
1	1.4759	0.5607	3.46%	-4.72%
2	1.4864	0.6211	4.51%	4.83%
3	1.4719	0.6284	3.06%	5.02%
4	1.3510	0.6215	-9.04%	-3.53%
5	1.3768	0.6314	-6.46%	-0.56%
6	1.3738	0.6130	-6.75%	-3.28%
7	1.4398	0.5902	-0.16%	-2.52%
8	1.4395	0.6059	-0.19%	-0.27%
9	1.4369	0.6129	-0.45%	0.58%
10	1.4351	0.6040	-0.62%	-0.81%
11	1.4852	0.6115	4.39%	3.34%
12	1.4735	0.5924	3.21%	-0.20%
13	1.5188	0.5512	7.74%	-3.76%
14	1.4634	0.5953	2.21%	-0.37%
15	1.4473	0.6240	0.59%	2.82%

divisi per 100.

Osserviamo che la media $\bar{f}_0^P = 1.4413$ è la stima della catena del fattore di sviluppo f_0^P . La media dei $(P/I)_{i,0}$ è $\overline{(P/I)}_{i,0} = 0.6075$.

Dai valori riportati nella tabella vi è un indizio di correlazione negativa tra i fattori $f_{i,0}^P$ e i rapporti $(P/I)_{i,0}$. Infatti a scarti positivi per $f_{i,0}^P$ corrispondono, spesso, scarti negativi per $(P/I)_{i,0}$ e viceversa. Calcolando il coefficiente di correlazione empirico

$$\hat{\rho} = \frac{\sum_{i=0}^{I-1} (f_{i,0}^P - \bar{f}_0^P) ((P/I)_{i,0} - \overline{(P/I)}_{i,0})}{\sqrt{\sum_{i=0}^{I-1} (f_{i,0}^P - \bar{f}_0^P)^2} \sqrt{\sum_{i=0}^{I-1} ((P/I)_{i,0} - \overline{(P/I)}_{i,0})^2}},$$

si ottiene -59.7%. Ciò può essere evidenziato disegnando un grafico avente sulle ascisse i rapporti $(P/I)_{i,0}$ e sulle ordinate i fattori $f_{i,0}^P$ (Figura 3.1): vi è chiara presenza di un trend decrescente. La retta orizzontale indica il valore medio dei fattori di sviluppo \bar{f}_0^P , mentre la retta verticale il valore medio dei rapporti (P/I) , $\overline{(P/I)}_{i,0}$.

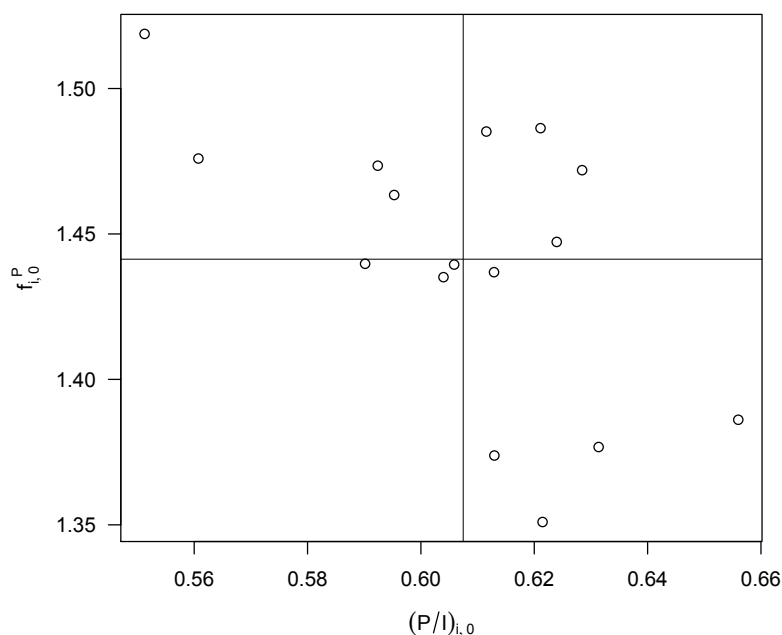


Figura 3.1: fattori di sviluppo $f_{i,0}^P$ e rapporti $(P/I)_{i,0}$.

Calcoliamo ora i fattori di sviluppo individuali per gli *incurred losses*

$$f_{i,0}^I = \frac{I_{i,1}}{I_{i,0}}, \quad i = 0, \dots, I - 1,$$

e gli scarti dalle rispettive medie divisi per 100

$$\Delta_3 = f_{i,0}^I - \bar{f}_0^I.$$

Nella Tabella 3.2 sono riportati i risultati. Il coefficiente di correlazione empirico tra tali fattori di sviluppo e i rapporti $(P/I)_{i,0}$ è 36.04%. Abbiamo perciò una situazione opposta alla precedente, con debole correlazione positiva. Come in precedenza costruiamo un grafico avente sulle ascisse i rapporti $(P/I)_{i,0}$ e sulle ordinate i fattori $f_{i,0}^I$ (Figura 3.2).

Le correlazioni evidenziate hanno il seguente significato: generalmente, a un rapporto $(P/I)_{i,0}$ sopra la media corrisponde un fattore di sviluppo $f_{i,0}^P$ sotto la media e un fattore di sviluppo $f_{i,0}^I$ sopra la media. Viceversa per un rapporto $(P/I)_{i,0}$ sotto la media. Questo è evidenziato nella Tabella 3.3, dove sono confrontati gli scarti dalla media dei rapporti $(P/I)_{i,0}$ e dei fattori di sviluppo individuali $f_{i,0}^P$ e $f_{i,0}^I$. \square

L'analisi sviluppata nell'esempio riguarda i rapporti (P/I) relativi al differimento 0 e i fattori di sviluppo individuali relativi al passaggio tra il differimento 0 e il differimento 1.

Tabella 3.2: confronto tra fattori di sviluppo individuali $f_{i,0}^I$ e rapporti $(P/I)_{i,0}$.

i	$f_{i,0}^I$	$(P/I)_{i,0}$	$\Delta_3\%$	$\Delta_2\%$
0	1.0005	0.9093	4.85%	3.44%
1	0.9510	0.8276	-4.67%	-4.72%
2	1.0552	0.9232	1.36%	4.83%
3	1.0323	0.9250	2.10%	5.02%
4	0.9621	0.8396	1.40%	-3.53%
5	0.9818	0.8692	2.39%	-0.56%
6	0.9577	0.8421	0.55%	-3.28%
7	0.9589	0.8497	-1.73%	-2.52%
8	0.9464	0.8721	-0.16%	-0.27%
9	0.9769	0.8806	0.54%	0.58%
10	0.9899	0.8668	-0.35%	-0.81%
11	1.0238	0.9083	0.41%	3.34%
12	1.0090	0.8729	-1.51%	-0.20%
13	0.9828	0.8372	-5.62%	-3.76%
14	1.0021	0.8711	-1.22%	-0.37%
15	1.0184	0.9030	1.65%	2.82%

Tuttavia, essa lascia intendere che una situazione analoga si potrà presentare al variare del differimento.

Questo dovrà essere tenuto in considerazione nella proiezione dei triangoli di run-off, poiché non sarà corretto usare per ogni anno di origine lo stesso fattore di sviluppo medio f_j^P e f_j^I .

3.2 Idee preliminari per la soluzione del problema (P/I)

Al fine di stimare fattori di sviluppo adeguati, intuitivamente, potremmo pensare di stimare una retta di regressione per i punti in ciascuna delle figure 3.1 e 3.2, passante per il punto con coordinate i valori medi, come mostrato in Figura 3.3. Potremmo poi usare il fattore di sviluppo previsto dalla retta di regressione, anziché il valor medio, per i diversi anni di origine. In tale modo, i fattori di sviluppo sono maggiori o minori della media a

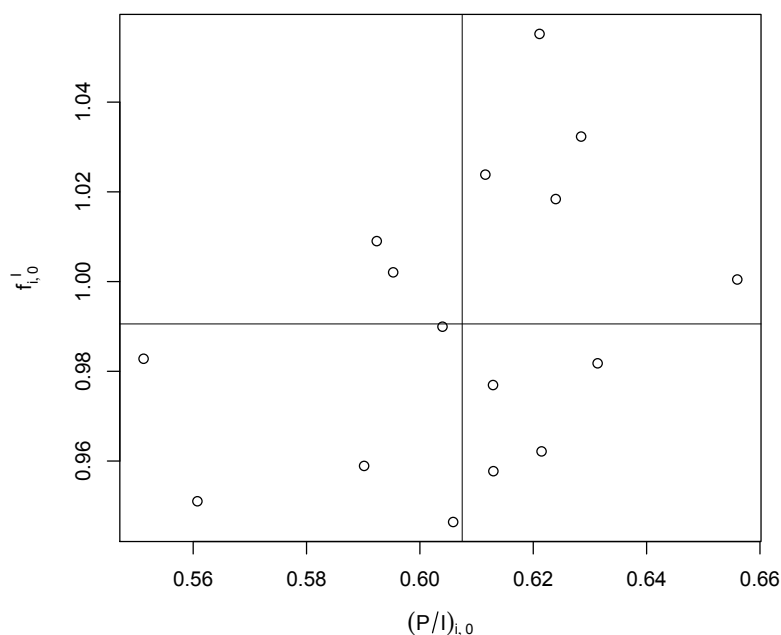


Figura 3.2: fattori di sviluppo $f_{i,0}^I$ e rapporti $(P/I)_{i,0}$.

seconda del valore del rapporto (P/I) corrispondente. Questa analisi andrebbe ripetuta per ogni anno di sviluppo.

Tale metodo presenta però una criticità che verrà evidenziata nell'esempio seguente.

Esempio. Consideriamo i dati relativi a un portafoglio RCA di Generali US Branch¹. I triangoli di run-off hanno dimensione $I = 9$. In Figura 3.4 viene mostrata la retta di regressione stimata per i fattori di sviluppo individuali $f_{i,0}^P$ e i rapporti $(P/I)_{i,0}$; dall'analisi è stato escluso l'anno di origine $i = 1$ in quanto outlier ($f_{1,0}^P = 338$).

Tale modellizzazione non è buona: la retta non realizza un buon accostamento ai dati. Inoltre, si stimerebbe per l'anno di origine $i = 9$ un fattore di sviluppo negativo, mentre i fattori di sviluppo devono essere positivi. Una curva iperbolica sarebbe stata più adatta a descrivere i dati. Ponendo infatti sulle ascisse i rapporti (I/P) , anziché i rapporti (P/I) , otteniamo il grafico in Figura 3.6: ora la retta di regressione è adatta a descrivere i dati.

Riportiamo ora in Figura 3.5 la retta di regressione stimata per i fattori di sviluppo individuali $f_{i,0}^I$ e i rapporti $(P/I)_{i,0}$. In questo caso la criticità evidenziata sopra non è presente. Osserviamo che con questi dati otteniamo un coefficiente di correlazione empirico

¹Fonte http://www.casact.org/research/index.cfm?fa=loss_reserves_data.

Tabella 3.3: scarti dalla media dei rapporti $(P/I)_{i,0}$ e dei fattori di sviluppo individuali $f_{i,0}^P$ e $f_{i,0}^I$.

i	$\Delta_2\%$	$\Delta_1\%$	$\Delta_3\%$
0	3.44%	-5.52%	4.85%
1	-4.72%	3.46%	-4.67%
2	4.83%	4.51%	1.36%
3	5.02%	3.06%	2.10%
4	-3.53%	-9.04%	1.40%
5	-0.56%	-6.46%	2.39%
6	-3.28%	-6.75%	0.55%
7	-2.52%	-0.16%	-1.73%
8	-0.27%	-0.19%	-0.16%
9	0.58%	-0.45%	0.54%
10	-0.81%	-0.62%	-0.35%
11	3.34%	4.39%	0.41%
12	-0.20%	3.21%	-1.51%
13	-3.76%	7.74%	-5.62%
14	-0.37%	2.21%	-1.22%
15	2.82%	0.59%	1.65%

negativo tra i fattori $f_{i,0}^I$ e i rapporti $(P/I)_{i,0}$, pari a -0.7875, mentre generalmente tale coefficiente risulta positivo.

□

L'esempio ha mostrato che vi possono essere situazioni nelle quali considerare la retta di regressione per i punti $((P/I)_{i,0}, f_{i,0}^P)$ non fornisce un buon modello per stimare i fattori di sviluppo. Con i dati dell'esempio e più in generale, se i dati presentano un andamento iperbolico, una soluzione è il passaggio ai reciproci dei rapporti (P/I) , ovvero ai rapporti (I/P) . In tal caso i punti $((P/I)_{i,0}, f_{i,0}^P)$ si dispongono in modo tale da essere ben approssimati da una retta. Nel caso dei dati di Figura 3.1, che sono bene descritti dalla retta di regressione, il passaggio ai reciproci dei valori in ascissa mantiene tale andamento, come mostrato in Figura 3.7.

Possiamo quindi pensare di legare i fattori di sviluppo $f_{i,0}^P$ ai rapporti $(I/P)_{i,0}$. Per i

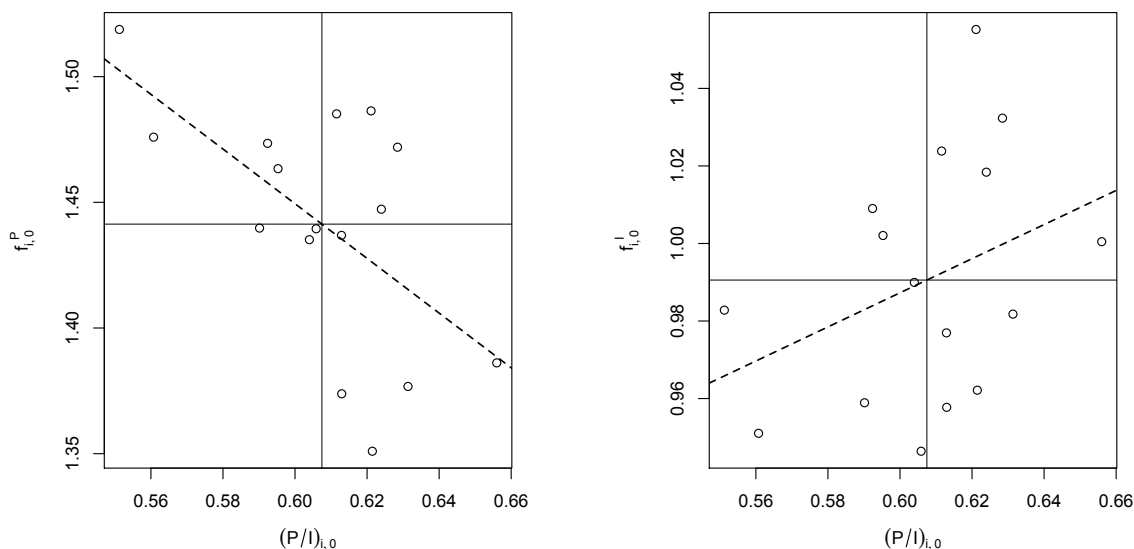


Figura 3.3: rette di regressione per i fattori di sviluppo.

fattori $f_{i,0}^I$ possiamo invece mantenere il legame f con i rapporti $(P/I)_{i,0}$. L'analisi andrebbe ripetuta per gli altri differimenti.

Un ulteriore problema che emerge però nella pratica è la scarsità di dati a disposizione per la stima della retta di regressione per anni di sviluppo elevati. L'idea per superare tale difficoltà è considerare la totalità degli anni di sviluppo per la stima della retta, e non il singolo anno di sviluppo. Per fare ciò occorre standardizzare i rapporti (P/I) e (I/P) e i fattori di sviluppo $f_{i,j}^P$ e $f_{i,j}^I$, per far sì che gli ordini di grandezza siano confrontabili; questo viene fatto attraverso il passaggio ai residui standardizzati di tali valori. Precisamente, consideriamo la seguente standardizzazione. Dati y_1, \dots, y_n valori osservati di una grandezza di interesse, indicati con \bar{y} la media e σ_y lo scarto quadratico medio empirico, poniamo

$$r_h = \frac{y_h - \bar{y}}{\sigma_y}, \quad h = 1, \dots, n.$$

Esempio (continua). In Figura 3.8 abbiamo considerato i dati delle figure 3.7 e 3.2 e abbiamo effettuato il passaggio ai residui dei fattori di sviluppo e dei rapporti (I/P) e (P/I) .

Possiamo ora stimare le rette di regressione per i punti dei due grafici e utilizzarle per il calcolo dei fattori di sviluppo per ogni anno di origine e per ogni anno di sviluppo. Si ottiene in entrambi i casi la retta $y = x$. \square

Il procedimento seguito nell'esempio può essere sintetizzato come segue:

- (i) si considerano le coppie dei residui relativi ai rapporti $(I/P)_{i,j}$ e ai fattori di sviluppo individuali dei pagamenti $f_{i,j}^P$, per ogni i, j ;

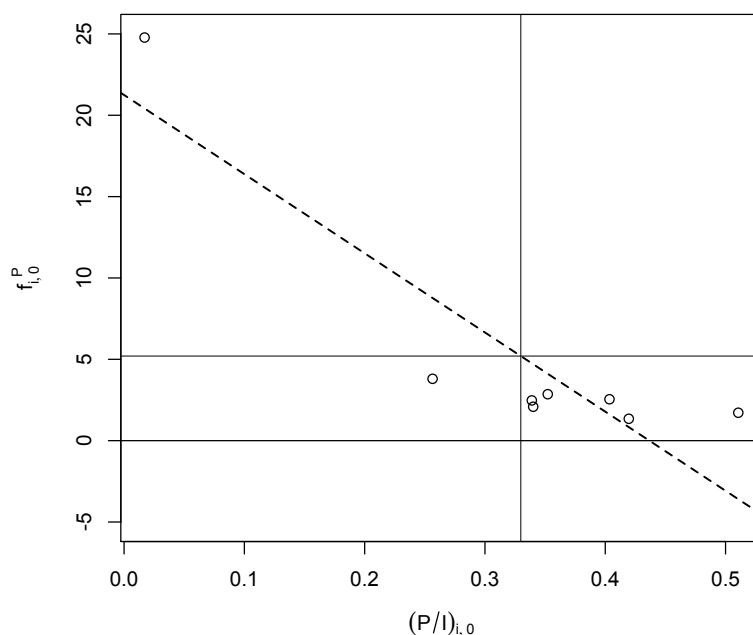


Figura 3.4: retta di regressione per i fattori di sviluppo individuali.

- (ii) si considerano le coppie dei residui relativi ai rapporti $(P/I)_{i,j}$ e ai fattori di sviluppo individuali degli *incurred losses* $f_{i,j}^I$, per ogni i, j ;
- (iii) si calcolano le rette di regressione passanti per l'origine, per entrambi i tipi di dati;
- (iv) con riferimento ai pagamenti cumulati, a fronte di un rapporto $(P/I)_{i,j}$: si considera il residuo; si legge il residuo del fattore di sviluppo associato dalla retta di regressione; da tale residuo si calcola il fattore di sviluppo $f_{i,j}^P$. Tale fattore di sviluppo sostituisce quello che si otterrebbe con il metodo della catena.
- (v) si eseguono passi analoghi a quelli descritti al punto (iv) con riferimento agli *incurred losses*.

Tali considerazioni euristiche verranno formalizzate nel prossimo paragrafo nelle ipotesi del metodo MCL.

3.3 Ipotesi del modello

Siano $\{P_{i,j}, j = 0, \dots, I\}$ e $\{I_{i,j}, j = 0, \dots, I\}$ i processi stocastici dei pagamenti cumulati e degli *incurred losses* per l'anno di origine i . Indichiamo con $\mathcal{P}_i(j) = \{P_{i,0}, \dots, P_{i,j}\}$ e $\mathcal{I}_i(j) = \{I_{i,0}, \dots, I_{i,j}\}$ l'informazione disponibile per i pagamenti e gli *incurred losses*

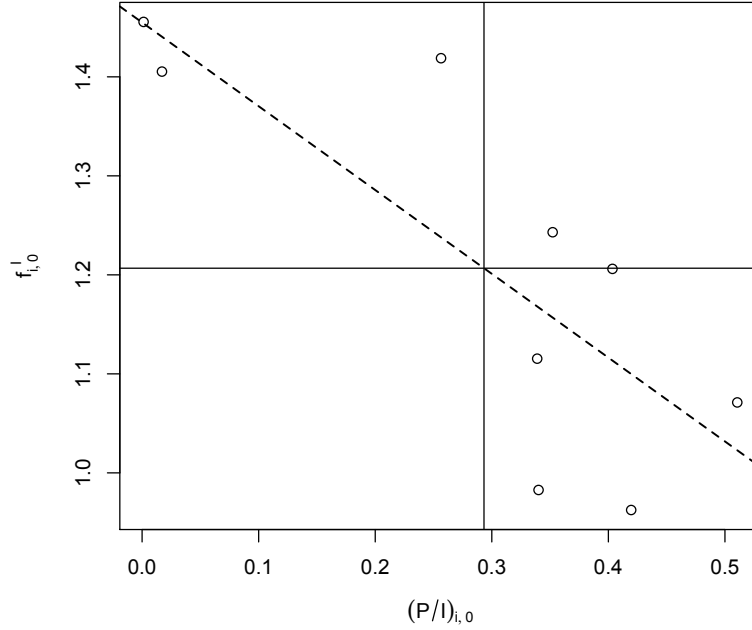


Figura 3.5: retta di regressione per i fattori di sviluppo individuali.

dell'anno di origine i fino all'anno di sviluppo j . Sia inoltre $\mathcal{B}_i(j) = \{P_{i,0}, \dots, P_{i,j}, I_{i,0}, \dots, I_{i,j}\}$ l'informazione congiunta disponibile fino all'anno di sviluppo j , per l'anno di origine i . Indichiamo inoltre con $\mathcal{P}(j) = \{P_{i,k}, k \leq j, 0 \leq i \leq I\}$, $\mathcal{I}(j) = \{I_{i,k}, k \leq j, 0 \leq i \leq I\}$, $\mathcal{B}(j) = \{P_{i,k}, I_{i,k}, k \leq j, 0 \leq i \leq I\}$.

Poniamo

$$Q_{i,j} := \frac{P_{i,j}}{I_{i,j}}$$

il rapporto tra i pagamenti cumulati e gli *incurred losses* con riferimento all'anno di origine i e all'anno di sviluppo j .

Ipotesi 3.1 (MCL).

- (i) $\{P_{0,j}, I_{0,j}, j = 0, \dots, I\}, \dots, \{P_{I,j}, I_{I,j}, j = 0, \dots, I\}$ sono stocasticamente indipendenti.
- (ii) Esistono $f_0^P, \dots, f_{I-1}^P > 0$ tali che

$$\begin{aligned} \mathbf{E}(P_{i,j+1} | \mathcal{P}(j)) &= \mathbf{E}(P_{i,j+1} | \mathcal{P}_i(j)) \\ &= \mathbf{E}(P_{i,j+1} | P_{i,j}) = P_{i,j} \cdot f_j^P, \quad j = 0, \dots, I-1, \quad i = 0, \dots, I. \end{aligned} \quad (3.1)$$

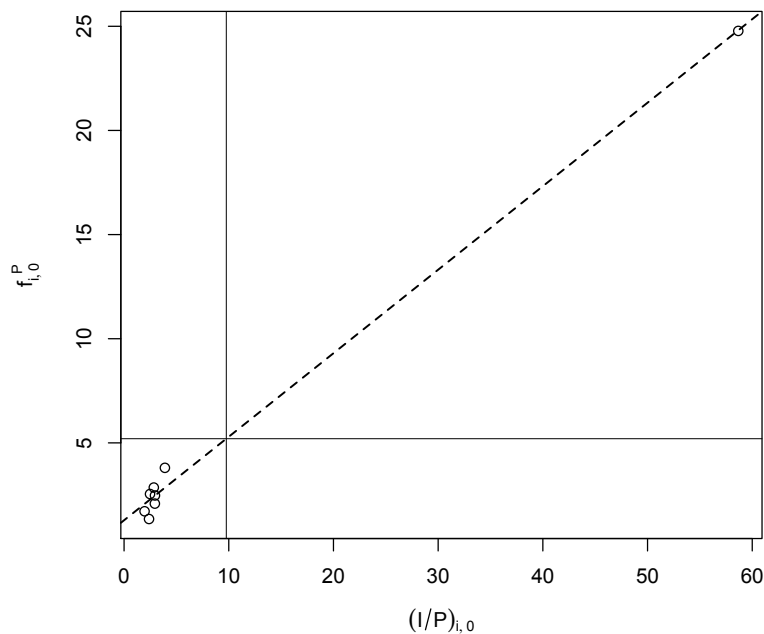


Figura 3.6: retta di regressione per i fattori di sviluppo.

(iii) Esistono $\sigma_0^P, \dots, \sigma_{I-1}^P > 0$ tali che

$$\begin{aligned} \mathbf{Var}(P_{i,j+1}|\mathcal{P}(j)) &= \mathbf{Var}(P_{i,j+1}|\mathcal{P}_i(j)) \\ &= \mathbf{Var}(P_{i,j+1}|P_{i,j}) = P_{i,j} \cdot (\sigma_j^P)^2, \quad j = 0, \dots, I-1, \quad i = 0, \dots, I. \end{aligned} \quad (3.2)$$

(iv) Esistono $f_0^I, \dots, f_{I-1}^I > 0$ tali che

$$\begin{aligned} \mathbf{E}(I_{i,j+1}|\mathcal{I}(j)) &= \mathbf{E}(I_{i,j+1}|\mathcal{I}_i(j)) \\ &= \mathbf{E}(I_{i,j+1}|I_{i,j}) = I_{i,j} \cdot f_j^I, \quad j = 0, \dots, I-1, \quad i = 0, \dots, I. \end{aligned} \quad (3.3)$$

(v) Esistono $\sigma_0^I, \dots, \sigma_{I-1}^I > 0$ tali che

²La prima uguaglianza è conseguenza dell'indipendenza dei pagamenti relativi a diversi anni di origine.

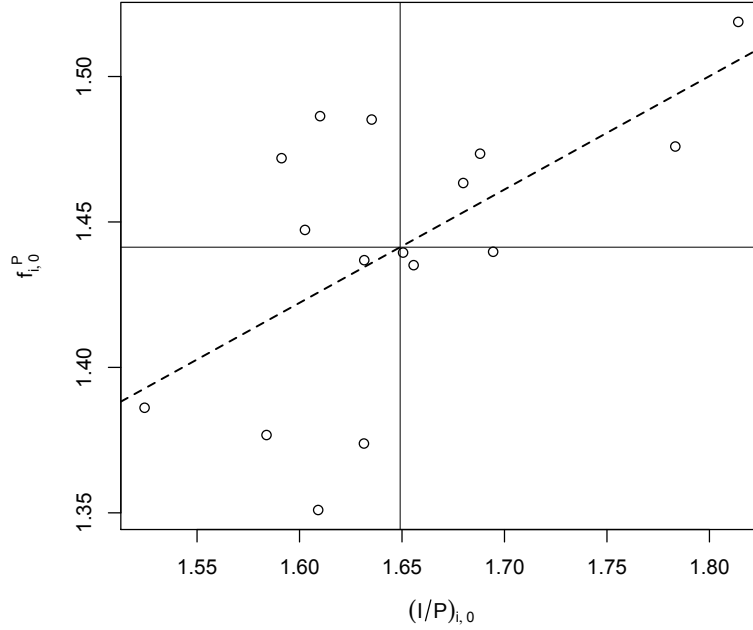


Figura 3.7: passaggio ai rapporti (I/P) per i dati in Figura 3.1.

$$\begin{aligned}
 \mathbf{Var}(I_{i,j+1}|\mathcal{I}(j)) &= \mathbf{Var}(I_{i,j+1}|\mathcal{I}_i(j)) \\
 &= \mathbf{Var}(I_{i,j+1}|I_{i,j}) = I_{i,j} \cdot (\sigma_j^I)^2, \quad j = 0, \dots, I-1, \quad i = 0, \dots, I.
 \end{aligned}
 \tag{3.4}$$

(vi) Esiste una costante λ^P tale che

$$\begin{aligned}
 \mathbf{E}\left(\frac{P_{i,j+1}}{P_{i,j}}|\mathcal{B}(j)\right) &= \mathbf{E}\left(\frac{P_{i,j+1}}{P_{i,j}}|\mathcal{B}_i(j)\right) \\
 &= f_j^P + \lambda^P \cdot \mathbf{Var}\left(\frac{P_{i,j+1}}{P_{i,j}}|\mathcal{P}_i(j)\right)^{1/2} \cdot \frac{(Q_{i,j}^{-1} - \mathbf{E}(Q_{i,j}^{-1}|\mathcal{P}_i(j)))}{\mathbf{Var}(Q_{i,j}^{-1}|\mathcal{P}_i(j))^{1/2}}
 \end{aligned}
 \tag{3.5}$$

per $j = 0, \dots, I-1$ e $i = 0, \dots, I$.

(vii) Esiste una costante λ^I tale che

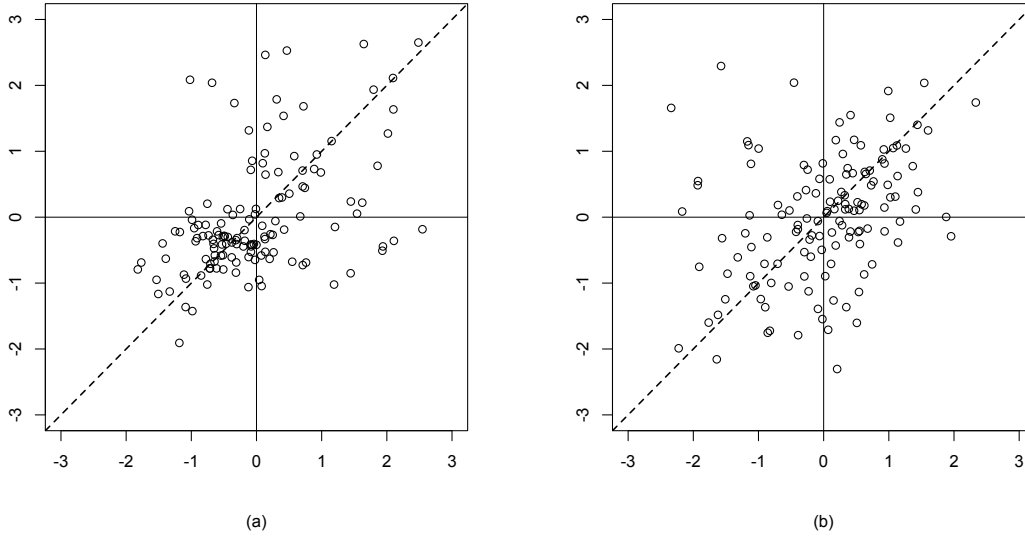


Figura 3.8: sulle ascisse i residui dei rapporti (I/P) e (P/I) e sulle ordinate i residui dei fattori di sviluppo individuali $f_{i,j}^P$ e $f_{i,j}^I$.

$$\begin{aligned} \mathbf{E}\left(\frac{I_{i,j+1}}{I_{i,j}} \mid \mathcal{B}(j)\right) &= \mathbf{E}\left(\frac{I_{i,j+1}}{I_{i,j}} \mid \mathcal{B}_i(j)\right) \\ &= f_j^I + \lambda^I \cdot \mathbf{Var}\left(\frac{I_{i,j+1}}{I_{i,j}} \mid \mathcal{I}_i(j)\right)^{1/2} \cdot \frac{(Q_{i,j} - \mathbf{E}(Q_{i,j} \mid \mathcal{I}_i(j)))}{\mathbf{Var}(Q_{i,j} \mid \mathcal{I}_i(j))^{1/2}} \end{aligned} \quad (3.6)$$

per $j = 0, \dots, I-1$ e $i = 0, \dots, I$. □

Osserviamo che l'ipotesi (i) è una condizione di indipendenza più forte rispetto a supporre separatamente

$\{P_{0,j}, j = 0, \dots, I\}, \dots, \{P_{I,j}, j = 0, \dots, I\}$ stocasticamente indipendenti

e

$\{I_{0,j}, j = 0, \dots, I\}, \dots, \{I_{I,j}, j = 0, \dots, I\}$ stocasticamente indipendenti.

Le ipotesi da (i) a (v) implicano dunque quelle del modello di Mack, sia per i pagamenti sia per gli *incurred losses*. Le ipotesi (vi) e (vii) sono le assunzioni fondamentali del metodo MCL. Nel prossimo paragrafo presentiamo una analisi approfondita di tali ipotesi al fine di mettere in luce i collegamenti con le considerazioni sviluppate nei paragrafi 3.1 e 3.2.

3.4 Analisi delle ipotesi

Ci soffermiamo dapprima sull'ipotesi (vii) del MCL. Osserviamo che la condizione (3.6) richiede che il valore atteso condizionato $\mathbf{E}\left(\frac{I_{i,j+1}}{I_{i,j}}|\mathcal{B}_i(j)\right)$ sia una funzione lineare del rapporto $Q_{i,j} = (P/I)_{i,j}$. Precisamente, tale valore atteso è la somma del fattore di sviluppo f_j^I , che è l'usuale fattore di sviluppo del modello CL, e di un termine di correzione lineare in $Q_{i,j}$. In particolare, se $\lambda^I \geq 0$, la funzione lineare è non decrescente. L'ipotesi esprime formalmente l'andamento dei fattori di sviluppo individuali $f_{i,j}^I$ rispetto ai rapporti $(P/I)_{i,j}$ illustrato nella Figura 3.2 nel caso $j = 0$.

Analoghe considerazioni possono essere fatte per la condizione (3.5) dell'ipotesi (vi): il valore atteso condizionato $\mathbf{E}\left(\frac{P_{i,j+1}}{P_{i,j}}|\mathcal{B}_i(j)\right)$ è in questo caso una funzione lineare, monotona non decrescente se $\lambda^P \geq 0$, del rapporto $Q_{i,j}^{-1} = (I/P)_{i,j}$. L'ipotesi traduce l'andamento dei fattori di sviluppo individuali $f_{i,j}^P$ rispetto ai rapporti $(I/P)_{i,j}$ illustrato nella Figura 3.7.

Analizziamo ora in maggior dettaglio i termini di correzione nelle (vi), (vii). A tal fine introduciamo il concetto di residuo condizionato che consentirà una diversa formulazione delle ipotesi (vi) e (vii) del MCL.

Definizione 3.1 (Residuo condizionato). Sia X una variabile aleatoria e \mathcal{C} una σ -algebra, il residuo condizionato di X data \mathcal{C} è

$$\mathbf{Res}(X|\mathcal{C}) := \frac{X - \mathbf{E}(X|\mathcal{C})}{\mathbf{Var}(X|\mathcal{C})^{1/2}} .$$

Il residuo condizionato è standardizzato rispetto alla speranza matematica condizionata e alla varianza condizionata, nel senso che sussistono le seguenti proprietà, di immediata verifica,

$$\mathbf{E}(\mathbf{Res}(X|\mathcal{C})|\mathcal{C}) = 0 \quad \text{e} \quad \mathbf{Var}(\mathbf{Res}(X|\mathcal{C})|\mathcal{C}) = 1.$$

La formula (3.5) si può allora scrivere come

$$\begin{aligned} \frac{\mathbf{E}\left(\frac{P_{i,j+1}}{P_{i,j}}|\mathcal{B}_i(j)\right) - \mathbf{E}\left(\frac{P_{i,j+1}}{P_{i,j}}|\mathcal{P}_i(j)\right)}{\mathbf{Var}\left(\frac{P_{i,j+1}}{P_{i,j}}|\mathcal{P}_i(j)\right)^{1/2}} &= \lambda^P \cdot \frac{\left(Q_{i,j}^{-1} - \mathbf{E}(Q_{i,j}^{-1}|\mathcal{P}_i(j))\right)}{\mathbf{Var}\left(Q_{i,j}^{-1}|\mathcal{P}_i(j)\right)^{1/2}}, \\ \mathbf{E}\left(\frac{\frac{P_{i,j+1}}{P_{i,j}} - \mathbf{E}\left(\frac{P_{i,j+1}}{P_{i,j}}|\mathcal{P}_i(j)\right)}{\mathbf{Var}\left(\frac{P_{i,j+1}}{P_{i,j}}|\mathcal{P}_i(j)\right)^{1/2}}\middle|\mathcal{B}_i(j)\right) &= \lambda^P \cdot \frac{\left(Q_{i,j}^{-1} - \mathbf{E}(Q_{i,j}^{-1}|\mathcal{P}_i(j))\right)}{\mathbf{Var}\left(Q_{i,j}^{-1}|\mathcal{P}_i(j)\right)^{1/2}}, \\ \mathbf{E}\left(\mathbf{Res}\left(\frac{P_{i,j+1}}{P_{i,j}}|\mathcal{P}_i(j)\right)\middle|\mathcal{B}_i(j)\right) &= \lambda^P \cdot \mathbf{Res}(Q_{i,j}^{-1}|\mathcal{P}_i(j)), \end{aligned}$$

essendo $\mathbf{E}\left(\frac{P_{i,j+1}}{P_{i,j}}|\mathcal{P}_i(j)\right)$ e $\mathbf{Var}\left(\frac{P_{i,j+1}}{P_{i,j}}|\mathcal{P}_i(j)\right)^{1/2}$ $\mathcal{P}_i(j)$ -misurabili e $\mathcal{B}_i(j)$ una sovra σ -algebra di $\mathcal{P}_i(j)$. Analogamente, la formula (3.6) diventa

$$\mathbf{E}\left(\mathbf{Res}\left(\frac{I_{i,j+1}}{I_{i,j}}|\mathcal{I}_i(j)\right)|\mathcal{B}_i(j)\right) = \lambda^I \cdot \mathbf{Res}(Q_{i,j}|\mathcal{I}_i(j)). \quad (3.7)$$

Tale formulazione traduce in termini teorici il legame lineare tra i residui suggerito dagli esempi nei paragrafi 3.1 e 3.2. Si riveda in particolare la Figura 3.8.

Riprendiamo il termine di correzione della (3.6) e presentiamo una interpretazione dei tre fattori che lo costituiscono.

- λ^I è il coefficiente di correlazione, supposto indipendente da i e j , tra i residui condizionati dei fattori di sviluppo individuali $f_{i,j}^I = I_{i,j+1}/I_{i,j}$ e i residui condizionati dei rapporti $(P/I)_{i,j} = Q_{i,j}$. Infatti, si ha

$$\begin{aligned} \mathbf{Cov}\left(Q_{i,j}, \frac{I_{i,j+1}}{I_{i,j}}|\mathcal{I}_i(j)\right) &= \mathbf{Cov}\left(Q_{i,j}, \mathbf{E}\left(\frac{I_{i,j+1}}{I_{i,j}}|\mathcal{B}_i(j)\right)|\mathcal{I}_i(j)\right) = \\ &= \mathbf{Cov}\left(Q_{i,j}, f_j^I + \lambda^I \cdot \frac{\mathbf{Var}\left(\frac{I_{i,j+1}}{I_{i,j}}|\mathcal{I}_i(j)\right)^{1/2}}{\mathbf{Var}\left(Q_{i,j}|\mathcal{I}_i(j)\right)^{1/2}} \cdot \left(Q_{i,j} - \mathbf{E}\left(Q_{i,j}|\mathcal{I}_i(j)\right)\right)\middle|\mathcal{I}_i(j)\right) = \\ &= \lambda^I \cdot \frac{\mathbf{Var}\left(\frac{I_{i,j+1}}{I_{i,j}}|\mathcal{I}_i(j)\right)^{1/2}}{\mathbf{Var}\left(Q_{i,j}|\mathcal{I}_i(j)\right)^{1/2}} \cdot \mathbf{Var}\left(Q_{i,j}|\mathcal{I}_i(j)\right) = \\ &= \lambda^I \cdot \mathbf{Var}\left(\frac{I_{i,j+1}}{I_{i,j}}|\mathcal{I}_i(j)\right)^{1/2} \cdot \mathbf{Var}\left(Q_{i,j}|\mathcal{I}_i(j)\right)^{1/2}, \end{aligned}$$

dove la prima uguaglianza segue da una proprietà della covarianza condizionata³, essendo $Q_{i,j}$ $\mathcal{B}_i(j)$ -misurabile e $\mathcal{B}_i(j)$ una sovra σ -algebra di $\mathcal{I}_i(j)$, le altre dalla proprietà di linearità della covarianza e dal fatto che f_j^I , λ^I sono numeri certi.

Allora, riesce

$$\mathbf{Corr}\left(Q_{i,j}, \frac{I_{i,j+1}}{I_{i,j}}|\mathcal{I}_i(j)\right) = \frac{\mathbf{Cov}\left(Q_{i,j}, \frac{I_{i,j+1}}{I_{i,j}}|\mathcal{I}_i(j)\right)}{\sqrt{\mathbf{Var}\left(Q_{i,j}|\mathcal{I}_i(j)\right)\mathbf{Var}\left(\frac{I_{i,j+1}}{I_{i,j}}|\mathcal{I}_i(j)\right)}} = \lambda^I. \quad (3.8)$$

³Se \mathcal{H} è una sovra σ -algebra di una σ -algebra \mathcal{K} , se X e Y sono due numeri aleatori di speranza matematica finita e X è \mathcal{H} -misurabile, allora si ha

$$\begin{aligned} \mathbf{Cov}(X, Y|\mathcal{K}) &= \mathbf{E}(XY|\mathcal{K}) - \mathbf{E}(X|\mathcal{K})\mathbf{E}(Y|\mathcal{K}) = \mathbf{E}(\mathbf{E}(XY|\mathcal{H})|\mathcal{K}) - \mathbf{E}(X|\mathcal{K})\mathbf{E}(Y|\mathcal{K}) \\ &= \mathbf{E}(X\mathbf{E}(Y|\mathcal{H})|\mathcal{K}) - \mathbf{E}(X|\mathcal{K})\mathbf{E}(\mathbf{E}(Y|\mathcal{H})|\mathcal{K}) = \mathbf{Cov}(X, \mathbf{E}(Y|\mathcal{H})|\mathcal{K}), \end{aligned}$$

dove la penultima uguaglianza segue dalla proprietà iterativa della speranza matematica.

Pertanto λ^I è il coefficiente di correlazione condizionato a $\mathcal{I}_i(j)$ di $Q_{i,j}$ e $I_{i,j+1}/I_{i,j}$. Consideriamo ora il coefficiente di correlazione, condizionato a $\mathcal{I}_i(j)$, dei residui condizionati $\mathbf{Res}(Q_{i,j}|\mathcal{I}_i(j))$, $\mathbf{Res}\left(\frac{I_{i,j+1}}{I_{i,j}}|\mathcal{I}_i(j)\right)$. Si ha

$$\begin{aligned}
 & \text{Corr}\left(\mathbf{Res}\left(Q_{i,j}|\mathcal{I}_i(j)\right), \mathbf{Res}\left(\frac{I_{i,j+1}}{I_{i,j}}|\mathcal{I}_i(j)\right)\middle|\mathcal{I}_i(j)\right) = \\
 & = \text{Cov}\left(\mathbf{Res}\left(Q_{i,j}|\mathcal{I}_i(j)\right), \mathbf{Res}\left(\frac{I_{i,j+1}}{I_{i,j}}|\mathcal{I}_i(j)\right)\middle|\mathcal{I}_i(j)\right) = \\
 & = \mathbf{E}\left[\mathbf{Res}\left(Q_{i,j}|\mathcal{I}_i(j)\right)\mathbf{Res}\left(\frac{I_{i,j+1}}{I_{i,j}}|\mathcal{I}_i(j)\right)\middle|\mathcal{I}_i(j)\right] \\
 & - \mathbf{E}\left[\mathbf{Res}\left(Q_{i,j}|\mathcal{I}_i(j)\right)\middle|\mathcal{I}_i(j)\right]\mathbf{E}\left[\mathbf{Res}\left(\frac{I_{i,j+1}}{I_{i,j}}|\mathcal{I}_i(j)\right)\middle|\mathcal{I}_i(j)\right] = \\
 & = \mathbf{E}\left[\mathbf{Res}\left(Q_{i,j}|\mathcal{I}_i(j)\right)\mathbf{Res}\left(\frac{I_{i,j+1}}{I_{i,j}}|\mathcal{I}_i(j)\right)\middle|\mathcal{I}_i(j)\right] = \\
 & = \mathbf{E}\left[\left(\frac{Q_{i,j} - \mathbf{E}(Q_{i,j}|\mathcal{I}_i(j))}{\mathbf{Var}(Q_{i,j}|\mathcal{I}_i(j))^{1/2}}\right)\left(\frac{\frac{I_{i,j+1}}{I_{i,j}} - \mathbf{E}\left(\frac{I_{i,j+1}}{I_{i,j}}|\mathcal{I}_i(j)\right)}{\mathbf{Var}\left(\frac{I_{i,j+1}}{I_{i,j}}|\mathcal{I}_i(j)\right)^{1/2}}\right)\middle|\mathcal{I}_i(j)\right] = \\
 & = \frac{1}{\mathbf{Var}(Q_{i,j}|\mathcal{I}_i(j))^{1/2}\mathbf{Var}\left(\frac{I_{i,j+1}}{I_{i,j}}|\mathcal{I}_i(j)\right)^{1/2}} \\
 & \cdot \mathbf{E}\left[\left(Q_{i,j} - \mathbf{E}(Q_{i,j}|\mathcal{I}_i(j))\right)\left(\frac{I_{i,j+1}}{I_{i,j}} - \mathbf{E}\left(\frac{I_{i,j+1}}{I_{i,j}}|\mathcal{I}_i(j)\right)\right)\middle|\mathcal{I}_i(j)\right] = \\
 & = \frac{\text{Cov}\left(Q_{i,j}, \frac{I_{i,j+1}}{I_{i,j}}\middle|\mathcal{I}_i(j)\right)}{\mathbf{Var}(Q_{i,j}|\mathcal{I}_i(j))^{1/2}\mathbf{Var}\left(\frac{I_{i,j+1}}{I_{i,j}}|\mathcal{I}_i(j)\right)^{1/2}} = \text{Corr}\left(Q_{i,j}, \frac{I_{i,j+1}}{I_{i,j}}\middle|\mathcal{I}_i(j)\right) = \lambda^I,
 \end{aligned}$$

dove abbiamo usato, nell'ordine, le seguenti proprietà: le varianze condizionate dei residui sono unitarie, le speranze matematiche condizionate dei residui sono nulle, $\mathbf{Var}(Q_{i,j}|\mathcal{I}_i(j))^{1/2}$ e $\mathbf{Var}\left(\frac{I_{i,j+1}}{I_{i,j}}|\mathcal{I}_i(j)\right)^{1/2}$ sono $\mathcal{I}_i(j)$ -misurabili e la (3.8). Dunque, il parametro λ^I è il coefficiente di correlazione lineare condizionato tra i residui condizionati dei $Q_{i,j}$ e dei fattori di sviluppo individuali $f_{i,j}^I = I_{i,j+1}/I_{i,j}$. Esso può essere interpretato come il coefficiente angolare della retta di regressione nel grafico dei residui rappresentata dalla Figura 3.8 (b), per l'esempio sviluppato nel Paragrafo 3.2.

Dalla (3.8), λ^I fornisce una indicazione della dipendenza tra i fattori di sviluppo individuali e i rapporti (P/I). Se non vi è dipendenza, $\lambda^I = 0$. Allora, nella (3.6) il termine di correzione è nullo e il metodo si riduce al modello Chain-Ladder.

- Il rapporto $\frac{\text{Var}\left(\frac{I_{i,j+1}}{I_{i,j}}|\mathcal{I}_i(j)\right)^{1/2}}{\text{Var}\left(Q_{i,j}|\mathcal{I}_i(j)\right)^{1/2}}$ riscalda le deviazioni dei rapporti $Q_{i,j}$ dalla media in deviazioni dei fattori di sviluppo dalla media.

Maggiore è il numeratore, ossia la standard deviation condizionata del fattore di sviluppo, maggiore è la dispersione della distribuzione del fattore di sviluppo rispetto alla media, maggiore è il fattore correttivo; minore è il denominatore, ossia la standard deviation condizionata del rapporto $Q_{i,j}$, più improbabile sarà una deviazione significativa di $Q_{i,j}$ dalla media, e maggiore è il termine correttivo.

- Il termine $\left(Q_{i,j} - \mathbf{E}(Q_{i,j}|\mathcal{I}_i(j))\right)$ corregge il fattore di sviluppo f_j^I verso l'alto se il rapporto $Q_{i,j}$ è superiore al proprio valore atteso condizionato, viceversa verso il basso se è inferiore. Se si ha un rapporto $Q_{i,j}$ pari al suo valore atteso condizionato non si avrà correzione e il fattore di sviluppo sarà quello del metodo Chain-Ladder.

Analoghe considerazioni possono essere sviluppate per la condizione (3.5) dell'ipotesi (vi): il coefficiente λ^P è il coefficiente di correlazione tra i residui condizionati dei fattori di sviluppo dei pagamenti e i residui condizionati dei rapporti (I/P).

Dalle precedenti osservazioni si ha che il modello tiene conto delle analisi sviluppate negli esempi introduttivi, dove si era evidenziato un effetto di correlazione tra i fattori di sviluppo individuali relativi ai pagamenti cumulati (e agli *incurred losses*) e i rapporti I/P (P/I).

3.5 Stimatori per i parametri

Introduciamo ora degli stimatori per permettere l'applicazione del metodo proposto. Per la stima dei fattori di sviluppo f_j^P , f_j^I , $(\sigma_j^P)^2$ e $(\sigma_j^I)^2$ sono proposti gli stimatori usuali del metodo SCL:

$$\tilde{f}_j^P = \frac{\sum_{i=0}^{I-j-1} P_{i,j+1}}{\sum_{i=0}^{I-j-1} P_{i,j}} \quad \text{e} \quad \tilde{f}_j^I = \frac{\sum_{i=0}^{I-j-1} I_{i,j+1}}{\sum_{i=0}^{I-j-1} I_{i,j}}, \quad j = 0, \dots, I-1,$$

$$(\tilde{\sigma}_j^P)^2 = \frac{1}{I-j-1} \sum_{i=0}^{I-j-1} P_{i,j} \left(\frac{P_{i,j+1}}{P_{i,j}} - \tilde{f}_j^P \right)^2 \quad \text{e}$$

$$(\tilde{\sigma}_j^I)^2 = \frac{1}{I-j-1} \sum_{i=0}^{I-j-1} I_{i,j} \left(\frac{I_{i,j+1}}{I_{i,j}} - \tilde{f}_j^I \right)^2, \quad j = 0, \dots, I-2,$$

$$\tilde{\sigma}_{I-1}^P = \tilde{\sigma}_{I-2}^P \quad \text{e} \quad \tilde{\sigma}_{I-1}^I = \tilde{\sigma}_{I-2}^I.$$

Occorre ora stimare i parametri λ^P e λ^I . Possiamo sfruttare le (3.4) e (3.7) che, essendo $\mathbf{Res}(Q_{i,j}|\mathcal{I}_i(j))$ e $\mathbf{Res}(Q_{i,j}^{-1}|\mathcal{P}_i(j))$ $\mathcal{B}_i(j)$ -misurabili e usando le (ii)-(iv), consentono di scrivere le seguenti relazioni

$$\lambda^P = \mathbf{E} \left(\frac{\frac{P_{i,j+1}}{P_{i,j}} - f_j^P}{\sqrt{(\sigma_j^P)^2 P_{i,j}}} \middle/ \frac{Q_{i,j}^{-1} - \mathbf{E}(Q_{i,j}^{-1}|\mathcal{P}_i(j))}{\mathbf{Var}(Q_{i,j}^{-1}|\mathcal{P}_i(j))^{1/2}} \middle| \mathcal{B}_i(j) \right) \quad (3.9)$$

$$\lambda^I = \mathbf{E} \left(\frac{\frac{I_{i,j+1}}{I_{i,j}} - f_j^I}{\sqrt{(\sigma_j^I)^2 I_{i,j}}} \middle/ \frac{Q_{i,j} - \mathbf{E}(Q_{i,j}|\mathcal{I}_i(j))}{\mathbf{Var}(Q_{i,j}|\mathcal{I}_i(j))^{1/2}} \middle| \mathcal{B}_i(j) \right). \quad (3.10)$$

Per stimare i parametri λ^P , λ^I cerchiamo stimatori per i residui condizionati $\mathbf{Res}(Q_{i,j}|\mathcal{I}_i(j))$, $\mathbf{Res}(Q_{i,j}^{-1}|\mathcal{P}_i(j))$, $\mathbf{Res}(\frac{P_{i,j+1}}{P_{i,j}}|\mathcal{P}_i(j))$ e $\mathbf{Res}(\frac{I_{i,j+1}}{I_{i,j}}|\mathcal{I}_i(j))$. Presentiamo dapprima stimatori per $\mathbf{E}(Q_{i,j}|\mathcal{I}_i(j))$, $\mathbf{Var}(Q_{i,j}|\mathcal{I}_i(j))$, $\mathbf{E}(Q_{i,j}^{-1}|\mathcal{P}_i(j))$ e $\mathbf{Var}(Q_{i,j}^{-1}|\mathcal{P}_i(j))$.

Per semplicità, verrà assunto che i valori attesi condizionati $\mathbf{E}(Q_{i,j}|\mathcal{I}_i(j))$ e $\mathbf{E}(Q_{i,j}^{-1}|\mathcal{P}_i(j))$ siano certi (cioè non dipendano da $\mathcal{I}_i(j)$ e $\mathcal{P}_i(j)$ rispettivamente) e costanti in i . Questa ipotesi è analoga a quella fatta per $\mathbf{E}(\frac{P_{i,j+1}}{P_{i,j}}|\mathcal{P}_i(j))$ e $\mathbf{E}(\frac{I_{i,j+1}}{I_{i,j}}|\mathcal{I}_i(j))$ in (3.1) e (3.3), rispettivamente.

Tale ipotesi permette di ricavare i seguenti stimatori

Lemma 3.1. *Nell'ipotesi che $\mathbf{E}(Q_{i,j}|\mathcal{I}(j)) = \mathbf{E}(Q_{i,j}|\mathcal{I}_i(j))$, $\mathbf{E}(Q_{i,j}^{-1}|\mathcal{P}(j)) = \mathbf{E}(Q_{i,j}^{-1}|\mathcal{P}_i(j))$ siano certi e non dipendano da i ,*

(a) *condizionatamente a $\mathcal{I}(j)$,*

$$\tilde{q}_j = \frac{1}{\sum_{i=0}^{I-j} I_{i,j}} \sum_{i=0}^{I-j} I_{i,j} \cdot Q_{i,j} = \frac{\sum_{i=0}^{I-j} P_{i,j}}{\sum_{i=0}^{I-j} I_{i,j}} \quad (3.11)$$

è stimatore non distorto di $q_j = \mathbf{E}(Q_{i,j}|\mathcal{I}(j)) = \mathbf{E}(Q_{i,j}|\mathcal{I}_i(j))$, $j = 0, \dots, I$.

(b) *condizionatamente a $\mathcal{P}(j)$,*

$$\tilde{q}_j^{-1} = \frac{1}{\sum_{i=0}^{I-j} P_{i,j}} \sum_{i=0}^{I-j} P_{i,j} \cdot Q_{i,j}^{-1} = \frac{\sum_{i=0}^{I-j} I_{i,j}}{\sum_{i=0}^{I-j} P_{i,j}} \quad (3.12)$$

è stimatore non distorto di $q_j^{-1} = \mathbf{E}(Q_{i,j}^{-1}|\mathcal{P}(j)) = \mathbf{E}(Q_{i,j}^{-1}|\mathcal{P}_i(j))$, $j = 0, \dots, I$.

Dimostrazione. Abbiamo che

$$\mathbf{E}(\tilde{q}_j|\mathcal{I}(j)) = \mathbf{E} \left(\frac{\sum_{i=0}^{I-j} P_{i,j}}{\sum_{i=0}^{I-j} I_{i,j}} \middle| \mathcal{I}(j) \right) = \frac{1}{\sum_{i=0}^{I-j} I_{i,j}} \sum_{i=0}^{I-j} \mathbf{E}(P_{i,j}|\mathcal{I}(j)) =$$

$$\begin{aligned}
 &= \frac{1}{\sum_{i=0}^{I-j} I_{i,j}} \sum_{i=0}^{I-j} \mathbf{E}(I_{i,j} Q_{i,j} | \mathcal{I}(j)) = \frac{1}{\sum_{i=0}^{I-j} I_{i,j}} \sum_{i=0}^{I-j} I_{i,j} \mathbf{E}(Q_{i,j} | \mathcal{I}(j)) = \\
 &= \mathbf{E}(Q_{i,j} | \mathcal{I}(j)) = q_j,
 \end{aligned}$$

avendo usato, nell'ordine, le seguenti proprietà: $\sum_{i=0}^{I-j} I_{i,j}$ e $I_{i,j}$ sono $\mathcal{I}(j)$ -misurabili, $\mathbf{E}(Q_{i,j} | \mathcal{I}(j))$ è costante al variare di i . La prova per il caso (b) è analoga. \square

Osservazione. L'ipotesi che $\mathbf{E}(Q_{i,j} | \mathcal{I}_i(j))$ e $\mathbf{E}(Q_{i,j}^{-1} | \mathcal{P}_i(j))$ siano costanti in i e certi, non è in accordo con quanto si osserva nella pratica. Perciò tale ipotesi non è adeguata. Essa però permette di giustificare la scelta degli stimatori (3.11) e (3.12).

Al fine di ottenere stimatori per le varianze condizionate di $Q_{i,j}$ e $Q_{i,j}^{-1}$, osserviamo che le ipotesi (3.2) e (3.4) si possono scrivere come

$$\mathbf{Var}\left(\frac{P_{i,j+1}}{P_{i,j}} | \mathcal{P}_i(j)\right) = \frac{(\sigma_j^P)^2}{P_{i,j}} \quad \text{e} \quad \mathbf{Var}\left(\frac{I_{i,j+1}}{I_{i,j}} | \mathcal{I}_i(j)\right) = \frac{(\sigma_j^I)^2}{I_{i,j}}.$$

In analogia con tale scrittura, Quarg e Mack (2004) assumono che gli stimatori di $\mathbf{Var}(Q_{i,j}^{-1} | \mathcal{P}_i(j))$ e $\mathbf{Var}(Q_{i,j} | \mathcal{I}_i(j))$ dipendano in maniera inversamente proporzionale da $P_{i,j}$ e $I_{i,j}$, rispettivamente. Pongono dunque

$$\widetilde{\mathbf{Var}}(Q_{i,j}^{-1} | \mathcal{P}_i(j)) = \frac{1}{P_{i,j}} \cdot \frac{1}{I-j} \sum_{h=0}^{I-j} P_{h,j} \cdot (Q_{h,j}^{-1} - \tilde{q}_j^{-1})^2, \quad j = 0, \dots, I-1, \quad (3.13)$$

$$\widetilde{\mathbf{Var}}(Q_{i,j} | \mathcal{I}_i(j)) = \frac{1}{I_{i,j}} \cdot \frac{1}{I-j} \sum_{h=0}^{I-j} I_{h,j} \cdot (Q_{h,j} - \tilde{q}_j)^2, \quad j = 0, \dots, I-1. \quad (3.14)$$

In Wüthrich e Merz (2008) sono invece proposti gli stimatori del seguente lemma

Lemma 3.2. *Siano, per $0 \leq i \leq I$ e $0 \leq j \leq J$,*

$$\widetilde{\mathbf{Var}}(Q_{i,j} | \mathcal{I}(j)) = \left(\sum_{i=0}^{I-j} I_{i,j} - \frac{\sum_{i=0}^{I-j} I_{i,j}^2}{\sum_{i=0}^{I-j} I_{i,j}} \right)^{-1} \sum_{i=0}^{I-j} I_{i,j} (Q_{i,j} - \tilde{q}_j)^2$$

e

$$\widetilde{\mathbf{Var}}(Q_{i,j}^{-1} | \mathcal{P}(j)) = \left(\sum_{i=0}^{I-j} P_{i,j} - \frac{\sum_{i=0}^{I-j} P_{i,j}^2}{\sum_{i=0}^{I-j} P_{i,j}} \right)^{-1} \sum_{i=0}^{I-j} P_{i,j} (Q_{i,j}^{-1} - \tilde{q}_j^{-1})^2$$

gli stimatori per $\mathbf{Var}(Q_{i,j} | \mathcal{I}(j))$ e $\mathbf{Var}(Q_{i,j}^{-1} | \mathcal{P}(j))$, rispettivamente. Sotto le ipotesi del Lemma 3.1 e le ipotesi aggiuntive che $\mathbf{Var}(Q_{i,j} | \mathcal{I}(j))$, $\mathbf{Var}(Q_{i,j}^{-1} | \mathcal{P}(j))$, $\mathbf{Cov}(Q_{i,j}, \tilde{q}_j | \mathcal{I}(j))$ e $\mathbf{Cov}(Q_{i,j}^{-1}, \tilde{q}_j^{-1} | \mathcal{P}(j))$ siano costanti in i e certi, si ha che

(a) condizionatamente a $\mathcal{I}(j)$, $\widetilde{\mathbf{Var}}(Q_{i,j} | \mathcal{I}(j))$ è uno stimatore non distorto per $\mathbf{Var}(Q_{i,j} | \mathcal{I}(j))$.

(b) condizionatamente a $\mathcal{P}(j)$, $\widetilde{\mathbf{Var}}(Q_{i,j}^{-1}|\mathcal{P}(j))$ è uno stimatore non distorto per $\mathbf{Var}(Q_{i,j}^{-1}|\mathcal{P}(j))$.

Dimostrazione. Proviamo il caso (a). Essendo \tilde{q}_j , condizionatamente a $\mathcal{I}(j)$, uno stimatore non distorto per $\mathbf{E}(Q_{i,j}|\mathcal{I}(j))$, abbiamo la seguente relazione

$$\begin{aligned} \mathbf{E}[(Q_{i,j} - \tilde{q}_j)^2|\mathcal{I}(j)] &= \mathbf{E}[Q_{i,j}^2|\mathcal{I}(j)] + \mathbf{E}[\tilde{q}_j^2|\mathcal{I}(j)] - 2\mathbf{E}[Q_{i,j}\tilde{q}_j|\mathcal{I}(j)] = \\ &= (\mathbf{E}[Q_{i,j}^2|\mathcal{I}(j)] - \mathbf{E}[\tilde{q}_j|\mathcal{I}(j)]^2) + (\mathbf{E}[\tilde{q}_j^2|\mathcal{I}(j)] - \mathbf{E}[\tilde{q}_j|\mathcal{I}(j)]^2) + \\ &\quad - 2\mathbf{E}[Q_{i,j}\tilde{q}_j|\mathcal{I}(j)] + 2\mathbf{E}[\tilde{q}_j|\mathcal{I}(j)]^2 = \\ &= \mathbf{Var}(Q_{i,j}|\mathcal{I}(j)) + \mathbf{Var}(\tilde{q}_j|\mathcal{I}(j)) - 2\mathbf{Cov}(Q_{i,j}, \tilde{q}_j|\mathcal{I}(j)). \end{aligned}$$

Osserviamo inoltre che, essendo $I_{i,j}$ $\mathcal{I}(j)$ -misurabile e sfruttando l'indipendenza tra anni di origine,

$$\begin{aligned} \mathbf{Var}(\tilde{q}_j|\mathcal{I}(j)) &= \mathbf{Var}\left(\frac{\sum_{i=0}^{I-j} P_{i,j}}{\sum_{i=0}^{I-j} I_{i,j}} \middle| \mathcal{I}(j)\right) = \frac{1}{(\sum_{i=0}^{I-j} I_{i,j})^2} \mathbf{Var}\left(\sum_{i=0}^{I-j} P_{i,j} \middle| \mathcal{I}(j)\right) = \\ &= \frac{1}{(\sum_{i=0}^{I-j} I_{i,j})^2} \sum_{i=0}^{I-j} \mathbf{Var}(P_{i,j}|\mathcal{I}(j)) = \frac{1}{(\sum_{i=0}^{I-j} I_{i,j})^2} \sum_{i=0}^{I-j} I_{i,j}^2 \mathbf{Var}(Q_{i,j}|\mathcal{I}(j)) \end{aligned}$$

e

$$\begin{aligned} \mathbf{Cov}(Q_{i,j}, \tilde{q}_j|\mathcal{I}(j)) &= \mathbf{Cov}\left(\frac{P_{i,j}}{I_{i,j}}, \frac{\sum_{i=0}^{I-j} P_{i,j}}{\sum_{i=0}^{I-j} I_{i,j}} \middle| \mathcal{I}(j)\right) = \\ &= \frac{1}{I_{i,j} \sum_{i=0}^{I-j} I_{i,j}} \mathbf{Cov}\left(P_{i,j}, \sum_{i=0}^{I-j} P_{i,j} \middle| \mathcal{I}(j)\right) = \frac{1}{I_{i,j} \sum_{i=0}^{I-j} I_{i,j}} \sum_{i=0}^{I-j} \mathbf{Var}(P_{i,j}|\mathcal{I}(j)) = \\ &= \frac{I_{i,j}}{\sum_{i=0}^{I-j} I_{i,j}} \mathbf{Var}(Q_{i,j}|\mathcal{I}(j)). \end{aligned}$$

Abbiamo allora che

$$\begin{aligned} \mathbf{E}(\widetilde{\mathbf{Var}}(Q_{i,j}|\mathcal{I}(j))|\mathcal{I}(j)) &= \mathbf{E}\left(\left(\frac{\sum_{i=0}^{I-j} I_{i,j} - \frac{\sum_{i=0}^{I-j} I_{i,j}^2}{\sum_{i=0}^{I-j} I_{i,j}}\right)^{-1} \sum_{i=0}^{I-j} I_{i,j} (Q_{i,j} - \tilde{q}_j)^2 \middle| \mathcal{I}(j)\right) = \\ &= \left(\frac{\sum_{i=0}^{I-j} I_{i,j} - \frac{\sum_{i=0}^{I-j} I_{i,j}^2}{\sum_{i=0}^{I-j} I_{i,j}}\right)^{-1} \sum_{i=0}^{I-j} I_{i,j} \mathbf{E}\left((Q_{i,j} - \tilde{q}_j)^2 \middle| \mathcal{I}(j)\right) = \\ &= \left(\frac{\sum_{i=0}^{I-j} I_{i,j} - \frac{\sum_{i=0}^{I-j} I_{i,j}^2}{\sum_{i=0}^{I-j} I_{i,j}}\right)^{-1} \sum_{i=0}^{I-j} I_{i,j} \left(\mathbf{Var}(Q_{i,j}|\mathcal{I}(j)) + \frac{1}{(\sum_{i=0}^{I-j} I_{i,j})^2} \sum_{i=0}^{I-j} I_{i,j}^2 \mathbf{Var}(Q_{i,j}|\mathcal{I}(j)) + \right. \\ &\quad \left. - 2 \frac{I_{i,j}}{\sum_{i=0}^{I-j} I_{i,j}} \mathbf{Var}(Q_{i,j}|\mathcal{I}(j))\right) = \\ &= \mathbf{Var}(Q_{i,j}|\mathcal{I}(j)) \left(\frac{\sum_{i=0}^{I-j} I_{i,j} - \frac{\sum_{i=0}^{I-j} I_{i,j}^2}{\sum_{i=0}^{I-j} I_{i,j}}\right)^{-1} \left(\frac{\sum_{i=0}^{I-j} I_{i,j}}{\sum_{i=0}^{I-j} I_{i,j}} + \frac{\sum_{i=0}^{I-j} I_{i,j}^2}{\sum_{i=0}^{I-j} I_{i,j}} - 2 \frac{\sum_{i=0}^{I-j} I_{i,j}^2}{\sum_{i=0}^{I-j} I_{i,j}}\right) = \end{aligned}$$

$$= \mathbf{Var}(Q_{i,j}|\mathcal{I}(j)) \left(\sum_{i=0}^{I-j} I_{i,j} - \frac{\sum_{i=0}^{I-j} I_{i,j}^2}{\sum_{i=0}^{I-j} I_{i,j}} \right)^{-1} \left(\sum_{i=0}^{I-j} I_{i,j} - \frac{\sum_{i=0}^{I-j} I_{i,j}^2}{\sum_{i=0}^{I-j} I_{i,j}} \right) = \mathbf{Var}(Q_{i,j}|\mathcal{I}(j)).$$

La prova per il caso (b) è analoga. □

Nel seguito useremo gli stimatori di Quarg e Mack (3.13), (3.14).

Si ottengono i seguenti stimatori per i residui condizionati $\mathbf{Res}(Q_{i,j}|\mathcal{I}_i(j))$ e $\mathbf{Res}(Q_{i,j}^{-1}|\mathcal{P}_i(j))$:

$$\widetilde{\mathbf{Res}}(Q_{i,j}|\mathcal{I}_i(j)) = \frac{Q_{i,j} - \tilde{q}_j}{\sqrt{\widetilde{\mathbf{Var}}(Q_{i,j}|\mathcal{I}_i(j))}}$$

e

$$\widetilde{\mathbf{Res}}(Q_{i,j}^{-1}|\mathcal{P}_i(j)) = \frac{Q_{i,j}^{-1} - \tilde{q}_j^{-1}}{\sqrt{\widetilde{\mathbf{Var}}(Q_{i,j}^{-1}|\mathcal{P}_i(j))}}.$$

Per $\mathbf{Res}\left(\frac{P_{i,j+1}}{P_{i,j}}|\mathcal{P}_i(j)\right)$ e $\mathbf{Res}\left(\frac{I_{i,j+1}}{I_{i,j}}|\mathcal{I}_i(j)\right)$ sono proposti gli stimatori

$$\widetilde{\mathbf{Res}}\left(\frac{P_{i,j+1}}{P_{i,j}}|\mathcal{P}_i(j)\right) = \frac{\frac{P_{i,j+1}}{P_{i,j}} - \tilde{f}_j^P}{\sqrt{(\tilde{\sigma}_j^P)^2}} \cdot \sqrt{P_{i,j}}$$

e

$$\widetilde{\mathbf{Res}}\left(\frac{I_{i,j+1}}{I_{i,j}}|\mathcal{I}_i(j)\right) = \frac{\frac{I_{i,j+1}}{I_{i,j}} - \tilde{f}_j^I}{\sqrt{(\tilde{\sigma}_j^I)^2}} \cdot \sqrt{I_{i,j}}.$$

Possiamo ora stimare λ^P e λ^I . Le (3.9) e (3.10) suggeriscono i seguenti stimatori

$$\tilde{\lambda}^P = \frac{\sum_{0 \leq i+j \leq I} \widetilde{\mathbf{Res}}(Q_{i,j}^{-1}|\mathcal{P}_i(j)) \cdot \widetilde{\mathbf{Res}}\left(\frac{P_{i,j+1}}{P_{i,j}}|\mathcal{P}_i(j)\right)}{\sum_{0 \leq i+j \leq I} \widetilde{\mathbf{Res}}(Q_{i,j}^{-1}|\mathcal{P}_i(j))^2}$$

e

$$\tilde{\lambda}^I = \frac{\sum_{0 \leq i+j \leq I} \widetilde{\mathbf{Res}}(Q_{i,j}|\mathcal{I}_i(j)) \cdot \widetilde{\mathbf{Res}}\left(\frac{I_{i,j+1}}{I_{i,j}}|\mathcal{I}_i(j)\right)}{\sum_{0 \leq i+j \leq I} \widetilde{\mathbf{Res}}(Q_{i,j}|\mathcal{I}_i(j))^2}.$$

Dalle (3.5) e (3.6), otteniamo infine le seguenti formule ricorsive per la stima dei triangoli inferiori:

$$\tilde{P}_{i,j+1} = \tilde{P}_{i,j} \cdot \left(\tilde{f}_j^P + \tilde{\lambda}^P \cdot \frac{\tilde{\sigma}_j^P}{\sqrt{\tilde{I}_{i,j}}} \cdot \widetilde{\mathbf{Res}}(Q_{i,j}^{-1} | \mathcal{L}_i(j)) \right)$$

e

$$\tilde{I}_{i,j+1} = \tilde{I}_{i,j} \cdot \left(\tilde{f}_j^I + \tilde{\lambda}^I \cdot \frac{\tilde{\sigma}_j^I}{\sqrt{\tilde{P}_{i,j}}} \cdot \widetilde{\mathbf{Res}}(Q_{i,j} | \mathcal{P}_i(j)) \right),$$

con $i + j > I$ e valori iniziali $\tilde{P}_{i,I-i} = P_{i,I-i}$ e $\tilde{I}_{i,I-i} = I_{i,I-i}$.

Gli stimatori del costo ultimo per i diversi anni di origine saranno quindi

$$\tilde{P}_{i,I} \quad \text{e} \quad \tilde{I}_{i,I}, \quad i = 0, \dots, I,$$

mentre per il costo ultimo di portafoglio avremo

$$\sum_{i=0}^I \tilde{P}_{i,I} \quad \text{e} \quad \sum_{i=0}^I \tilde{I}_{i,I}.$$

Le riserve sinistri si possono ottenere infine con

$$\tilde{R}^P = \sum_{i=0}^I (\tilde{P}_{i,I} - P_{i,I-i}) \quad \text{e} \quad \tilde{R}^I = \sum_{i=0}^I (\tilde{I}_{i,I} - P_{i,I-i}).$$

3.6 Esemplicazioni numeriche

In questo paragrafo applichiamo il metodo MCL ad alcuni dataset, utilizzando il linguaggio R il pacchetto `ChainLadder` in particolare. Tale pacchetto verrà descritto in dettaglio nel Capitolo 6.

Esempio. Consideriamo i dati in Wüthrich e Merz (2008) e applichiamo gli stimatori definiti nel paragrafo precedente. In Tabella 3.4 sono riportate le stime dei parametri. I parametri $\hat{\sigma}_{15}^P$ e $\hat{\sigma}_{15}^I$ sono stati stimati per mezzo di una regressione log-lineare, ovvero stimando un modello lineare avente come variabile risposta $\log(\sigma_i^*)$, come regressori f_j^* , con $* \in \{P, I\}$, e facendo previsione con esso per $j = 15$.

Come è già stato osservato, le stime dei fattori di sviluppo e dei parametri di varianza coincidono con le usuali stime che si avrebbero applicando il metodo CL ai due triangoli di dati, separatamente. Per i coefficienti di correlazione otteniamo le stime $\hat{\lambda}^P = 0.4891$ e $\hat{\lambda}^I = 0.2388$; osserviamo che il coefficiente di correlazione empirico tra $f_{i,0}^P$ e $(I/P)_{i,0}$ è pari a 59.7% mentre quello tra $f_{i,0}^I$ e $(P/I)_{i,0}$ è pari a 36.04%, valori confrontabili con quelli ottenuti con il modello, ricordando però che i coefficienti di correlazione empirici sono stati calcolati utilizzando solo i dati del differimento $j = 0$.

Tabella 3.4: stime dei parametri.

j	\hat{f}_j^P	$\hat{\sigma}_j^P$	\hat{f}_j^I	$\hat{\sigma}_j^I$	\hat{q}_j	\hat{q}_j^{-1}
0	1.4416	11.8743	0.9903	9.9539	0.6066	1.6486
1	1.0278	5.6666	0.9787	5.6672	0.8826	1.1330
2	1.0112	2.5293	0.9868	3.5811	0.9266	1.0792
3	1.0057	1.4275	0.9888	4.0396	0.9488	1.0540
4	1.0048	1.3908	0.9917	3.2635	0.9655	1.0357
5	1.0025	1.3522	0.9993	2.2025	0.9773	1.0232
6	1.0008	0.5364	0.9986	1.3998	0.9801	1.0203
7	1.0020	1.6016	0.9973	0.6997	0.9823	1.0180
8	1.0010	0.6958	0.9990	0.3655	0.9868	1.0134
9	1.0001	0.0821	0.9986	0.4300	0.9887	1.0115
10	1.0002	0.1058	0.9986	0.4558	0.9899	1.0102
11	1.0019	0.5321	0.9967	3.0099	0.9897	1.0104
12	1.0002	0.0654	0.9999	0.0831	0.9951	1.0050
13	1.0000	0.0137	0.9995	0.1225	0.9954	1.0047
14	1.0001	0.0579	0.9998	0.0272	0.9956	1.0044
15	1.0001	(0.0236)	0.9990	(0.0620)	0.9936	1.0064
16					0.9892	1.0110

In Tabella 3.5 sono riportate le stime del costo ultimo, dei rapporti (P/I) finali e delle riserve per periodo di origine. Le stime delle riserve di portafoglio sono contenute in Tabella 3.6. Osserviamo che applicando il metodo SCL avremmo ottenuto come stime per le riserve di portafoglio 54,559 (*paid*) e 52,028 (*incurred*). Il metodo MCL ha corretto quindi verso il basso la stima *paid* e verso l'alto la stima *incurred*.

In Figura 3.9 è riportato infine un confronto tra il costo ultimo *paid* e il costo ultimo *incurred* per periodo di origine.

Come si vede dai rapporti (P/I) finali in Tabella 3.5 e dal grafico in Figura 3.9 le stime del costo ultimo ricavate utilizzando l'informazione sui pagamenti e l'informazione sugli *incurred losses* sono molto vicine. Diversamente, se si fosse applicato il metodo SCL, avremmo avuto scarti maggiori tra le stime dei costi ultimi *paid* e le stime dei costi ultimi *incurred*, come mostrato in Figura 3.10.

Tabella 3.5: stime del costo ultimo, dei rapporti (P/I) finali e delle riserve per periodo di origine.

i	$\hat{P}_{i,I}$	$\hat{I}_{i,I}$	$\hat{Q}_{i,I}\%$	\hat{R}_i^P	\hat{R}_i^I
0	67,336	68,074	98.92%	0	738
1	74,901	74,923	99.97%	3	25
2	86,293	86,234	100.07%	7	-52
3	95,692	95,910	99.77%	17	235
4	93,846	94,122	99.71%	35	311
5	94,068	94,383	99.67%	220	535
6	93,791	93,496	100.32%	89	-206
7	90,008	90,074	99.93%	176	242
8	93,108	93,240	99.86%	278	410
9	124,035	124,201	99.87%	589	755
10	113,702	113,927	99.80%	722	947
11	114,594	114,722	99.89%	789	917
12	103,854	103,899	99.96%	835	880
13	113,235	113,446	99.81%	2,335	2,546
14	102,887	102,990	99.90%	2,440	2,543
15	108,752	108,903	99.86%	5,700	5,851
16	110,421	110,577	99.86%	37,797	37,953

Tabella 3.6: stime delle riserve di portafoglio.

\hat{R}^P	\hat{R}^I
52,034	54,628

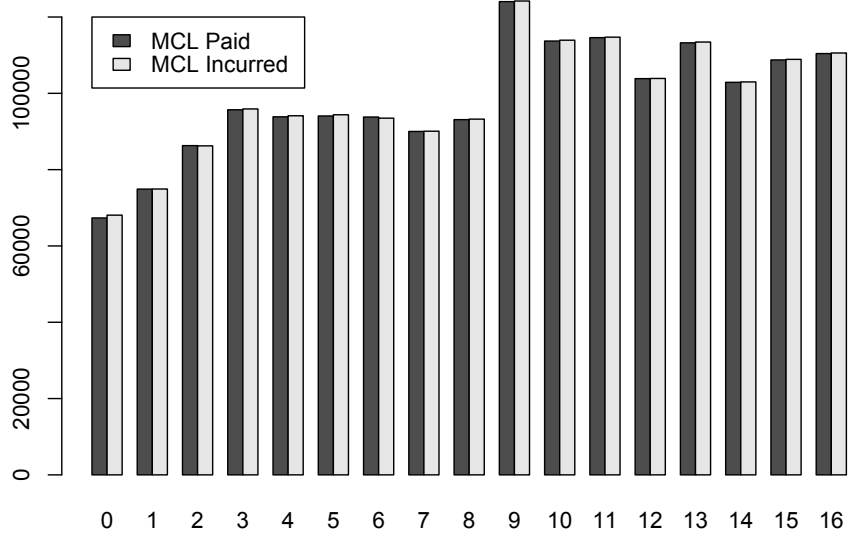


Figura 3.9: costo ultimo *paid* e *incurred* per periodo di origine.

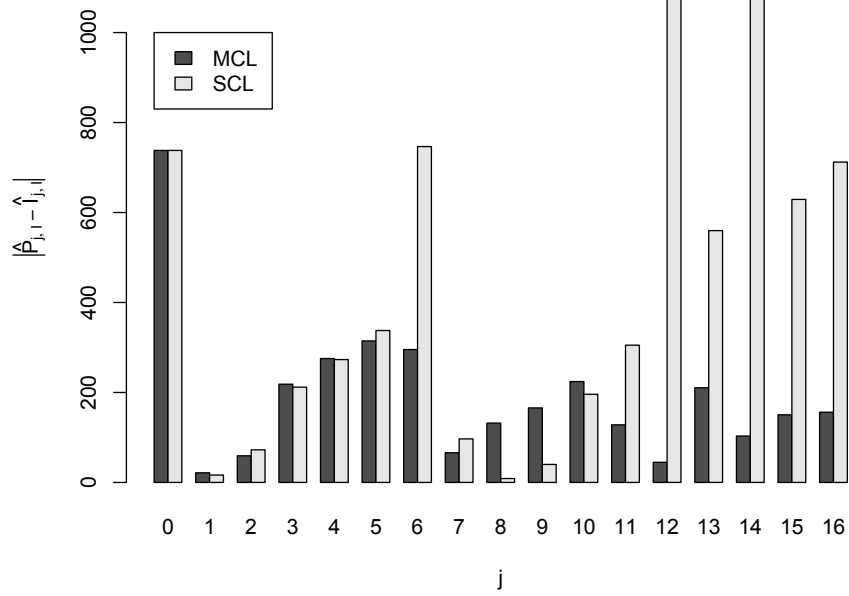


Figura 3.10: valore assoluto degli scarti tra costo ultimo *paid* e costo ultimo *incurred*.

Esempio. Consideriamo ora i dati del portafoglio della compagnia United Services Automobile Association già usato nell'esempio del Paragrafo 2.2. Nelle Tabelle 3.7 e 3.8 sono riportati i risultati per i costi ultimi e le riserve. Le stime dei coefficienti di correlazione sono $\hat{\lambda}^P = 0.6499$ e $\hat{\lambda}^I = -0.0105$. Ricordando l'interpretazione del parametro λ^I , riportato nel Paragrafo 3.4, abbiamo quindi correlazione quasi nulla tra i residui condizionati dei fattori di sviluppo degli *incurred losses* e i rapporti (P/I) . E' interessante confrontare le stime (P/I) finali ottenute con il metodo MCL con le stime che si sarebbero trovate applicando il metodo SCL. In Figura 3.11 è riportato tale confronto. Questo dataset bene evidenzia come il metodo MCL corregga le stime del metodo SCL, mantenendo i rapporti (P/I) finali più vicini a 100%.

Tabella 3.7: stime del costo ultimo e delle riserve per periodo di origine.

i	$\hat{P}_{i,I}$	$\hat{I}_{i,I}$	$\hat{Q}_{i,I}\%$	\hat{R}_i^P	\hat{R}_i^I
0	886,334	893,764	99.17%	0	7,430
1	982,823	985,293	99.75%	675	3,145
2	1,078,390	1,077,941	100.04%	2,853	2,404
3	1,143,969	1,135,464	100.75%	5,594	-2,911
4	1,240,551	1,226,776	101.12%	13,901	126
5	1,358,949	1,332,861	101.96%	34,217	8,129
6	1,402,252	1,366,620	102.61%	82,122	46,490
7	1,339,927	1,287,173	104.10%	154,627	101,873
8	1,278,702	1,218,166	104.97%	312,540	252,004
9	1,144,228	1,056,743	108.28%	602,207	514,722

Tabella 3.8: stime delle riserve di portafoglio.

\hat{R}^P	\hat{R}^I
1,208,737	933,410

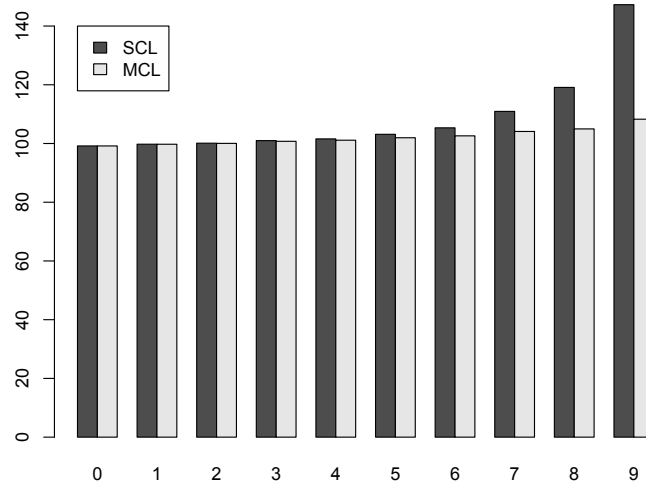


Figura 3.11: confronto delle stime dei rapporti (P/I) tra SCL e MCL.

Capitolo 4

Modello Paid-Incurred Chain

Il modello *Paid-Incurred Chain* (PIC) è stato presentato in Merz e Wüthrich (2010). Esso utilizza l'informazione sui pagamenti e sugli *incurred losses* per fornire una stima unica del costo ultimo, anziché due stime separate. Il modello sfrutta proprietà elementari della distribuzione Normale multidimensionale per unire i due canali di informazione. Il modello fornisce un peso di credibilità per ciascuna fonte d'informazione e permette inoltre di quantificare l'incertezza legata alla previsione.

4.1 Ipotesi del modello

Come in precedenza, siano $i \in \{0, \dots, I\}$ gli anni di origine e $j \in \{0, \dots, I\}$ gli anni di sviluppo. Le variabili aleatorie $P_{i,j}$ e $I_{i,j}$ indicano, rispettivamente, i pagamenti cumulati e gli *incurred losses* relativi all'anno di origine i e anno di sviluppo j . Assumiamo che

$$P_{i,I} = I_{i,I}, \quad i = 0, \dots, I.$$

Ipotizziamo quindi che i costi ultimi di pagamenti e *incurred losses* siano uguali per ogni anno di origine. Siano poi, per $j = 0, \dots, I$,

$$\mathcal{B}_j^P = \{P_{i,l} : 0 \leq i \leq I, 0 \leq l \leq j\},$$

$$\mathcal{B}_j^I = \{I_{i,l} : 0 \leq i \leq I, 0 \leq l \leq j\},$$

$$\mathcal{B}_j = \mathcal{B}_j^P \cup \mathcal{B}_j^I,$$

gli insiemi relativi all'informazione disponibile fino all'anno di sviluppo j sui pagamenti, sugli *incurred losses* e su entrambi i tipi di dati. Si assume per i pagamenti cumulati una formula ricorsiva del tipo

$$P_{i,j} = P_{i,j-1} \cdot \exp(\xi_{i,j}), \quad i = 0, \dots, I, \quad j = 1, \dots, I,$$

con $P_{i,0} = \exp(\xi_{i,j})$ e per gli *incurred losses* una formula ricorsiva all'indietro con valore iniziale $I_{i,I} = P_{i,I}$

$$I_{i,j-1} = I_{i,j} \cdot \exp(\zeta_{i,j-1}), \quad i = 0, \dots, I, \quad j = 1, \dots, I-1,$$

dove $\xi_{i,j}$ e $\zeta_{i,j}$ sono variabili aleatorie. Si noti che i pagamenti cumulati e gli *incurred losses* sono allora tutti positivi. Sia infine $\Theta = (\Phi_0, \dots, \Phi_I, \Psi_0, \dots, \Psi_{I-1}, \sigma_0, \dots, \sigma_I, \tau_0, \dots, \tau_{I-1})$ un vettore aleatorio di parametri.

Enunciamo ora le ipotesi del modello *Paid-Incurred Chain*.

Ipotesi 4.1 (PIC). • Condizionatamente a Θ , si ha

- il vettore aleatorio $(\xi_{0,0}, \dots, \xi_{I,I}, \zeta_{0,0}, \dots, \zeta_{I,I-1})$ ha distribuzione Normale multi-dimensionale con componenti indipendenti

$$\begin{aligned}\xi_{i,j} &\sim N(\Phi_j, \sigma_j^2), & i = 0, \dots, I, & \quad j = 0, \dots, I, \\ \zeta_{i,j} &\sim N(\Psi_j, \tau_j^2), & i = 0, \dots, I, & \quad j = 0, \dots, I-1;\end{aligned}$$

- i pagamenti cumulati $P_{i,j}$ sono dati dalla formula ricorsiva

$$P_{i,j} = P_{i,j-1} \cdot \exp(\xi_{i,j}), \quad (4.1)$$

con valore iniziale $P_{i,0} = \exp(\xi_{i,0})$;

- gli *incurred losses* $I_{i,j}$ sono dati dalla formula ricorsiva all'indietro

$$I_{i,j-1} = I_{i,j} \cdot \exp(-\zeta_{i,j-1}), \quad (4.2)$$

con valore iniziale $I_{i,I} = P_{i,I}$.

- Le componenti del vettore Θ sono indipendenti e $\sigma_j, \tau_j > 0$ per ogni j .

□

Osservazioni. L'assunzione che $P_{i,I} = I_{i,I}$ garantisce che i costi ultimi concidano utilizzando i dati sui pagamenti o i dati sugli *incurred losses*. Questo significa che, a differenza del modello Munich Chain Ladder, non si ha differenza tra le stime ottenute con i due canali di informazione.

L'assunzione che le componenti del vettore aleatorio $(\xi_{0,0}, \dots, \xi_{I,I}, \zeta_{0,0}, \dots, \zeta_{I,I-1})$ siano indipendenti potrebbe essere messa in discussione, dato che nel modello Munich Chain Ladder è stata evidenziata la presenza di correlazione tra i rapporti dei pagamenti sugli *incurred losses*. Tale dipendenza può in effetti essere implementata ma in tal caso le soluzioni si troverebbero solo per via numerica.

La scelta di un modello distribuzionale log-normale permette il calcolo esplicito delle distribuzioni condizionate di $P_{i,I}|\{\Theta, \mathcal{B}_j\}$, $P_{i,I}|\{\Theta, \mathcal{B}_j^P\}$ e $P_{i,I}|\{\Theta, \mathcal{B}_j^I\}$. Altre scelte distribuzionali richiederebbero metodi numerici.

4.2 Proprietà dei pagamenti cumulati e degli incurred losses

Per mezzo delle Ipotesi 4.1 ricaviamo ora delle espressioni per la valutazione dei costi ultimi. Osserviamo preliminarmente che, dalla (4.1), si ha

$$P_{i,j} = \prod_{m=0}^j \exp \xi_{i,m}. \quad (4.3)$$

Dunque, il pagamento cumulato $P_{i,j}$ è funzione dei numeri aleatori $(\xi_{i,0}, \dots, \xi_{i,j})$. Si ha inoltre che $\exp(\xi_{i,j})|\Theta$ ha distribuzione log-normale di parametri ϕ_j , σ_j e dunque

$$\mathbf{E}\left(\exp(\xi_{i,j})|\Theta\right) = \exp(\phi_j + \sigma_j^2/2). \quad (4.4)$$

Dalle Ipotesi 4.1 discende allora che poiché $P_{i,j-1}$ è \mathcal{B}_{j-1}^P -misurabile

$$\mathbf{E}\left(P_{i,j}|\{\Theta, \mathcal{B}_{j-1}^P\}\right) = P_{i,j-1} \cdot \mathbf{E}(\exp(\xi_{i,j})|\Theta, \mathcal{B}_{j-1}^P) = P_{i,j-1} \cdot \exp\{\Phi_j + \sigma_j^2/2\},$$

essendo $\exp(\xi_{i,j})|\Theta$ indipendente da \mathcal{B}_{j-1}^P per la (4.3), dall'indipendenza dei $\xi_{i,j}|\Theta$, e sfruttando la (4.4). Notiamo quindi che $\exp\{\Phi_j + \sigma_j^2/2\}$ può essere accostato al fattore di sviluppo per i pagamenti del metodo della catena. Abbiamo inoltre, sfruttando la ricorsione e la proprietà iterativa della speranza matematica condizionata

$$\begin{aligned} \mathbf{E}\left(P_{i,I}|\{\Theta, \mathcal{B}_j^P\}\right) &= \mathbf{E}\left(\mathbf{E}\left(P_{i,I}|\Theta, \mathcal{B}_j^P, P_{i,j+1}, \dots, P_{i,I-i}\right)|\{\Theta, \mathcal{B}_j^P\}\right) = \\ &= \mathbf{E}\left(P_{i,I-i} \exp(\phi_I + \sigma_I^2/2)|\{\Theta, \mathcal{B}_j^P\}\right) = \\ &= \exp(\phi_I + \sigma_I^2/2) \mathbf{E}(P_{i,I-i}|\{\Theta, \mathcal{B}_j^P\}) = \\ &= P_{i,j} \cdot \exp\left\{\sum_{l=j+1}^I (\Phi_l + \sigma_l^2/2)\right\}. \end{aligned} \quad (4.5)$$

La formula (4.5) è la speranza matematica del costo ultimo per l'anno di origine i , condizionatamente a Θ , sfruttando solo l'informazione sui pagamenti disponibile in $i + j$: \mathcal{B}_j^P .

Vogliamo ora ricavare una formula per la previsione dei costi ultimi che sfrutti solo l'informazione sugli *incurred losses*. Per farlo, richiamiamo preliminarmente una nota proprietà della distribuzione Normale multidimensionale, la quale verrà utilizzata per determinare la distribuzione degli *incurred losses*.

Lemma 4.1. *Sia $(X_1, \dots, X_n)'$ un vettore aleatorio con distribuzione Normale multidimensionale con valore atteso $(m_1, \dots, m_n)'$ e matrice di varianza-covarianza Σ , definita positiva. Abbiamo allora*

$$X_1|\{X_2, \dots, X_n\} \sim N\left(m_1 + \Sigma_{1,2}\Sigma_{2,2}^{-1}(X^{(2)} - m^{(2)}), \Sigma_{1,1} - \Sigma_{1,2}\Sigma_{2,2}^{-1}\Sigma_{2,1}\right),$$

dove $X^{(2)} = (X_2, \dots, X_n)'$ ha distribuzione $N(m^{(2)}, \Sigma_{2,2})$, con $m^{(2)} = (m_2, \dots, m_n)'$ e $\Sigma_{2,2}$ la sottomatrice di Σ corrispondente a $(X_2, \dots, X_n)'$, definita positiva, $\Sigma_{1,1}$ è la varianza di X_1 e $\Sigma_{1,2} = \Sigma'_{2,1}$ è il vettore di covarianza tra X_1 e $X^{(2)}$.

Osserviamo ora che, dalle (4.2) e (4.3), si ha

$$\begin{aligned} I_{i,j} &= P_{i,I} \cdot \prod_{n=j}^{I-1} \exp(-\zeta_{i,n}) \\ &= \prod_{m=0}^I \exp(\xi_{i,m}) \cdot \prod_{n=j}^{I-1} \exp(-\zeta_{i,n}), \quad i = 0, \dots, I, j = 0, \dots, I. \end{aligned} \quad (4.6)$$

Dunque $I_{i,j}$ è funzione dei numeri aleatori $\xi_{i,0}, \dots, \xi_{i,I}$ e $\zeta_{i,j}, \dots, \zeta_{i,I-1}$. Per mezzo del Lemma 4.1 otteniamo il seguente risultato:

Proposizione 4.1. *Sotto le Ipotesi 4.1 si ha che*

$$\log I_{i,j+l} | \{\Theta, \mathcal{B}_j^I\} \sim N\left(\mu_{j+l} + \frac{v_{j+l}^2}{v_j^2}(\log I_{i,j} - \mu_j), v_{j+l}^2 \left(1 - \frac{v_{j+l}^2}{v_j^2}\right)\right), \quad 0 \leq j+l \leq I$$

dove

$$\mu_j = \sum_{m=0}^I \Phi_m - \sum_{n=j}^{I-1} \Psi_n \quad e \quad v_j^2 = \sum_{m=0}^I \sigma_m^2 + \sum_{n=j}^{I-1} \tau_n^2.$$

Nota. Si noti che, con le usuali convenzioni,

$$\mu_I = \sum_{m=0}^I \phi_m \quad e \quad v_I^2 = \sum_{m=0}^I \sigma_m^2.$$

Dimostrazione. Consideriamo dapprima, condizionatamente solo a Θ , il vettore

$$(\log I_{i,j+l}, \log I_{i,j}, \dots, \log I_{i,0}) | \Theta. \quad (4.7)$$

Dalla (4.6),

$$\log I_{i,j} = \sum_{m=0}^I \xi_{i,m} + \sum_{n=j}^{I-1} (-\zeta_{i,n}), \quad (4.8)$$

pertanto le componenti di (4.7) sono combinazioni lineari di componenti del vettore $(\xi_{0,0}, \dots, \zeta_{I,I-1})$. Allora, una qualunque combinazione lineare delle componenti del vettore (4.6) è una combinazione lineare di componenti di $(\xi_{0,0}, \dots, \zeta_{I,I-1})$. Poiché quest'ultimo vettore condizionatamente a Θ ha distribuzione normale, ne segue che il vettore (4.7) ha distribuzione normale¹. Dalla equazione (4.8) ricaviamo allora la distribuzione di $\log I_{i,j} | \Theta$

$$\log I_{i,j} | \Theta \sim \mathcal{N}\left(\underbrace{\sum_{m=0}^I \Phi_m - \sum_{n=j}^{I-1} \Psi_n}_{\mu_j}, \underbrace{\sum_{m=0}^I \sigma_m^2 + \sum_{n=j}^{I-1} \tau_n^2}_{v_j^2}\right).$$

Infatti, dalla prima delle Ipotesi 4.1, ricaviamo

$$\begin{aligned} \mathbf{E}(\log I_{i,j} | \Theta) &= \sum_{m=0}^I \mathbf{E}(\xi_{i,m} | \Theta) - \sum_{n=j}^{I-1} \mathbf{E}(\eta_{i,n} | \Theta) = \\ &= \sum_{m=0}^I \phi_m - \sum_{n=j}^{I-1} \psi_n \equiv \mu_j, \end{aligned} \quad (4.9)$$

¹Ricordiamo che un vettore aleatorio $(X_1, \dots, X_n)'$ ha distribuzione normale multidimensionale se e solo se ogni sua combinazione lineare ha distribuzione normale.

$$\begin{aligned}
 \mathbf{Var}(\log I_{i,j}|\Theta) &= \sum_{m=0}^I \mathbf{Var}(\xi_{i,m}|\Theta) - \sum_{n=j}^{I-1} \mathbf{Var}(\zeta_{i,n}|\Theta) = \\
 &= \sum_{m=0}^I \sigma_j^2 - \sum_{n=j}^{I-1} \tau_n^2 \equiv v_j.
 \end{aligned} \tag{4.10}$$

Abbiamo allora che il vettore

$$(\log I_{i,j+l}, \log I_{i,j}, \log I_{i,j-1}, \dots, \log I_{i,0})' |_{\{\Theta\}}$$

ha distribuzione Normale multivariata con valore atteso $(\mu_{j+l}, \mu_j, \mu_{j-1}, \dots, \mu_0)$ e matrice di varianza covarianza Σ con elementi non diagonali dati da

$$\text{Cov}(\log I_{i,h}, \log I_{i,k}|\Theta) = v_h^2, \quad h > k \in \{j+l, j, j-1, \dots, 0\}.$$

Infatti, dalla (4.8), si ha

$$\begin{aligned}
 \text{Cov}(\log I_{i,h}, \log I_{i,k}|\Theta) &= \\
 &= \text{Cov}\left(\sum_{m=0}^I \xi_{i,m} + \sum_{n=h}^{I-1} (-\zeta_{i,n}), \sum_{s=0}^I \xi_{i,s} + \sum_{r=k}^{I-1} (-\zeta_{i,r})|\Theta\right) = \\
 &= \text{Cov}\left(\sum_{m=0}^I \xi_{i,m}, \sum_{s=0}^I \xi_{i,s}|\Theta\right) + \text{Cov}\left(\sum_{m=0}^I \xi_{i,m}, \sum_{r=k}^{I-1} (-\zeta_{i,r})|\Theta\right) + \\
 &+ \text{Cov}\left(\sum_{n=h}^{I-1} (-\zeta_{i,n}), \sum_{s=0}^I \xi_{i,s}|\Theta\right) + \text{Cov}\left(\sum_{n=h}^{I-1} (-\zeta_{i,n}), \sum_{r=k}^{I-1} (-\zeta_{i,r})|\Theta\right) = \\
 &= \sum_{m=0}^I \sum_{s=0}^I \text{Cov}(\xi_{i,m}, \xi_{i,s}|\Theta) - \sum_{m=0}^I \sum_{r=k}^{I-1} \text{Cov}(\xi_{i,m}, \zeta_{i,r}|\Theta) + \\
 &- \sum_{n=h}^{I-1} \sum_{s=0}^I \text{Cov}(\zeta_{i,n}, \xi_{i,s}|\Theta) + \sum_{n=h}^{I-1} \sum_{r=k}^{I-1} \text{Cov}(\zeta_{i,n}, \zeta_{i,r}|\Theta),
 \end{aligned}$$

dove abbiamo sfruttato la proprietà di linearità della covarianza rispetto a ciascuno dei suoi argomenti. Dalla prima delle Ipotesi 4.1 si ha

$$\begin{aligned}
 \text{Cov}(\xi_{i,m}, \xi_{i,s}|\Theta) &= \begin{cases} 0 & \text{se } s \neq m \\ \mathbf{Var}(\xi_{i,m}) & \text{se } s = m \end{cases}, \\
 \text{Cov}(\xi_{i,m}, \zeta_{i,s}|\Theta) &= 0 = \text{Cov}(\zeta_{i,n}, \xi_{i,s}|\Theta) \quad \text{per ogni } m, n, s, \\
 \text{Cov}(\zeta_{i,n}, \zeta_{i,s}|\Theta) &= \begin{cases} 0 & \text{se } s \neq n \\ \mathbf{Var}(\zeta_{i,n}) & \text{se } s = n \end{cases}.
 \end{aligned}$$

Pertanto, si ottiene

$$\text{Cov}(\log I_{i,h}, \log I_{i,k}|\Theta) = \sum_{m=0}^I \sigma_m^2 + \sum_{n=h}^{I-1} \tau_n^2 = v_h^2. \tag{4.11}$$

Ora, poiché condizionatamente a Θ gli *incurred losses* di diversi anni di origine sono indipendenti, dell'informazione portata da \mathcal{B}_j^I su $I_{i,j+l}$ è rilevante sulla valutazione probabilistica solo l'informazione portata dagli *incurred losses* dell'anno di origine i , o equivalentemente dei loro logaritmi, possiamo applicare il Lemma 4.1 prendendo

$$X_1 = \log I_{i,j+l} | \{\Theta, \mathcal{B}_j^I\} \quad \text{e} \quad X^{(2)} = (\log I_{i,j}, \log I_{i,j-1}, \dots, \log I_{i,0})' | \{\Theta\},$$

con parametri

$$\begin{aligned} m_1 &= \mathbf{E}(X_1) = \mu_{j+l}, \\ m^{(2)} &= \mathbf{E}(X^{(2)}) = (\mu_j, \dots, \mu_0), \\ \Sigma_{1,1} &= \mathbf{Var}(X_1) = v_{j+l}^2, \\ \Sigma_{2,2} &= \mathbf{Var}(X^{(2)}) = (v_{\max(j+1-n, j+1-m)}^2)_{1 \leq n, m \leq j+1}, \\ \Sigma_{1,2} &= (\mathbf{Cov}(\log I_{i,j+l}, \log I_{i,j}), \dots, \mathbf{Cov}(\log I_{i,j+l}, \log I_{i,0})) = (v_{j+l}^2, \dots, v_{j+l}^2) \in \mathbb{R}^{j+1}. \end{aligned}$$

Osserviamo infatti che $\Sigma_{2,2}$ è data da

$$\Sigma_{2,2} = \begin{bmatrix} v_j^2 & v_j^2 & v_j^2 & v_j^2 & \dots & v_j^2 \\ v_j^2 & v_{j-1}^2 & v_{j-1}^2 & v_{j-1}^2 & \dots & v_{j-1}^2 \\ v_j^2 & v_{j-1}^2 & v_{j-2}^2 & v_{j-2}^2 & \dots & v_{j-2}^2 \\ v_j^2 & v_{j-1}^2 & v_{j-2}^2 & v_{j-3}^2 & \dots & v_{j-3}^2 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ v_j^2 & v_{j-1}^2 & v_{j-2}^2 & v_{j-3}^2 & \dots & v_0^2 \end{bmatrix}.$$

Dunque $\log I_{i,j+l} | \{\Theta, \mathcal{B}_j^I\}$ ha distribuzione normale. Per determinare i valori dei parametri della distribuzione occorre calcolare l'inversa della matrice $\Sigma_{2,2}$. L'inversa ha la forma (i calcoli sono stati effettuati utilizzando il Symbolic Math Toolbox di MATLAB[®])

$$\Sigma_{2,2}^{-1} = (b_{n,m})_{1 \leq n, m \leq j+1},$$

con elementi diagonali

$$\begin{aligned} b_{1,1} &= \frac{v_{j-1}^2}{v_j^2(v_{j-1}^2 - v_j^2)}, \\ b_{n,n} &= \frac{v_{j-n}^2 - v_{j+2-n}^2}{(v_{j+1-n}^2 - v_{j+2-n}^2)(v_{j-n}^2 - v_{j+1-n}^2)}, \quad n \in \{2, \dots, j\}, \\ b_{j+1,j+1} &= \frac{1}{v_0^2 - v_1^2}, \end{aligned}$$

e con elementi non diagonali nulli ad eccezione della sopradiagonale

$$b_{n,n+1} = \frac{-1}{v_{j-n}^2 - v_{j+1-n}^2}, \quad n \in \{1, \dots, j\},$$

e della sottodiagonale simmetrica $b_{n,n-1}$, $n \in \{2, \dots, j+1\}$. Tale matrice ha la seguente proprietà

$$\Sigma_{1,2}\Sigma_{2,2}^{-1} = (v_{j+1}^2/v_j^2, 0, \dots, 0) \in \mathbb{R}^{j+1}.$$

Applicando ora il Lemma 4.1 si ha la tesi. \square

Dalla Proposizione 4.1 discende il seguente corollario, che fornisce la speranza matematica del costo ultimo per l'anno di origine i , condizionatamente a Θ , usando solo l'informazione sugli *incurred losses*: \mathcal{B}_{j-1}^I .

Corollario 4.1. *Sotto le ipotesi del metodo PIC abbiamo che*

$$\begin{aligned} \mathbf{E}\left(I_{i,I} \mid \{\Theta, \mathcal{B}_j^I\}\right) &= I_{i,j}^{1-\alpha_j} \cdot \exp\left\{(1-\alpha_j) \sum_{l=j}^{I-1} \Psi_l + \alpha_j(\mu_I + v_I^2/2)\right\} = \\ &= I_{i,j} \cdot \exp\left\{\sum_{l=j}^{I-1} (\Psi_l + \tau_l^2/2)\right\} \cdot \exp\left\{\alpha_j\left(\mu_j - \log I_{i,j} - \sum_{l=j}^{I-1} \tau_l^2/2\right)\right\}, \end{aligned} \quad (4.12)$$

con peso di credibilità

$$\alpha_j = 1 - \frac{v_I^2}{v_j^2} = \frac{1}{v_j^2} \sum_{l=j}^{I-1} \tau_l^2. \quad (4.13)$$

Dimostrazione. Dalla Proposizione 4.1 abbiamo che $I_{i,I} \mid \{\Theta, \mathcal{B}_j^I\}$ ha distribuzione log-normale con valore atteso

$$\begin{aligned} \mathbf{E}\left(I_{i,I} \mid \{\Theta, \mathcal{B}_j^I\}\right) &= \exp\left\{\mu_I + \frac{v_I^2}{v_j^2}(\log I_{i,j} - \mu_j) + \frac{v_I^2}{2}\left(1 - \frac{v_I^2}{v_j^2}\right)\right\} \\ &= \exp\{\mu_I + (1-\alpha_j)(\log I_{i,j} - \mu_j) + \alpha_j v_I^2/2\} \\ &= I_{i,j} \exp\left\{\mu_I - \mu_j\right\} \exp\{-\alpha_j(\log I_{i,j} - \mu_j) + \alpha_j v_I^2/2\} \\ &= I_{i,j} \exp\left\{\sum_{l=j}^{I-1} \Psi_l\right\} \exp\{-\alpha_j(\log I_{i,j} - \mu_j) + \alpha_j v_I^2/2\}, \end{aligned}$$

essendo

$$\mu_I - \mu_j = \sum_{m=0}^I \Phi_m - \sum_{m=0}^I \Phi_m + \sum_{n=j}^{I-1} \Psi_n = \sum_{n=j}^{I-1} \Psi_n.$$

Osserviamo ancora che $v_j^2 - v_I^2 = \sum_{l=j}^{I-1} \tau_l^2$ e dunque si ottiene la (4.13), da ciò segue

$$\alpha_j v_I^2 - \sum_{l=j}^{I-1} \tau_l^2 = \sum_{l=j}^{I-1} \tau_l^2 \left(\frac{v_I^2}{v_j^2} - 1\right) = -\alpha_j \sum_{l=j}^{I-1} \tau_l^2,$$

da cui segue la tesi. \square

Confrontando la (4.5) con la (4.12) si ha che, mentre i costi ultimi $P_{i,I}$ ottenuti con i pagamenti cumulati, condizionatamente a Θ e \mathcal{B}_j^P , soddisfano una condizione analoga a quella del modello CL, per i costi ultimi ottenuti con gli *incurred losses*, condizionatamente a Θ e \mathcal{B}_j^I , il prodotto dei fattori di sviluppo $\exp\{\sum_{l=j}^{I-1}(\Psi_l + \tau_l^2/2)\}$ viene moltiplicato per un fattore correttivo

$$\exp\left\{\alpha_j\left(\mu_j - \log I_{i,j} - \sum_{l=j}^{I-1} \tau_l^2/2\right)\right\},$$

il quale lo aggiusta tenendo conto dell'esperienza passata, con peso di credibilità α_j . Si noti che il peso di credibilità è tanto più grande quanto più piccolo è j . Infatti, si ha

$$v_j^2 = \sum_{m=0}^I \sigma_m^2 + \sum_{n=j}^{I-1} \tau_n^2 > \sum_{m=0}^I \sigma_m^2 + \sum_{n=j+1}^{I-1} \tau_n^2 = v_{j+1}^2.$$

Pertanto, si trova che

$$\alpha_j = 1 - \frac{v_I^2}{v_j^2} > 1 - \frac{v_I^2}{v_{j+1}^2} = \alpha_{j+1}.$$

Il fatto che nell'attuale modello gli *incurred losses* non soddisfino ad una ipotesi analoga a quella del CL, mostra che il modello non è della classe MCL.

Avendo ricavato delle espressioni per la valutazione dei costi ultimi utilizzando i due canali di informazione (*paid* e *incurred*) separatamente, il nostro obiettivo è ora quello di ottenere una previsione dei costi ultimi basandoci sull'informazione congiunta. Il seguente risultato fornisce la distribuzione dei costi ultimi data l'informazione congiunta.

Teorema 4.1. *Sotto le ipotesi del metodo PIC abbiamo la seguente distribuzione per il costo ultimo $P_{i,I} = I_{i,I}$, dato $\{\Theta, \mathcal{B}_j\}$, $0 \leq j < I$,*

$$\log P_{i,I} | \{\Theta, \mathcal{B}_j\} \sim N\left(\mu_I + (1 - \beta_j)(\log P_{i,j} - \eta_j) + \beta_j(\log I_{i,j} - \mu_j), (1 - \beta_j)(v_I^2 - w_j^2)\right),$$

dove

$$\eta_j = \sum_{m=0}^j \Phi_m \quad e \quad w_j^2 = \sum_{m=0}^j \sigma_m^2,$$

e il peso di credibilità è dato da

$$\beta_j = \frac{v_I^2 - w_j^2}{v_j^2 - w_j^2} > 0.$$

Dimostrazione. Analogamente a quanto fatto nella dimostrazione della Proposizione 4.1 consideriamo dapprima per l'anno di origine i , condizionatamente a Θ , il vettore

$$(\log I_{i,I}, \log P_{i,j}, \log I_{i,j}, \dots, \log I_{i,0}) | \Theta. \quad (4.14)$$

Poiché le componenti del vettore sono combinazioni lineari di elementi del vettore $(\xi_{0,0}, \dots, \zeta_{I,I-1})$, considerazioni analoghe a quelle sviluppate per la Proposizione ... consentono di concludere

che (4.14) ha distribuzione normale multidimensionale. Le speranze matematiche e le varianze delle componenti del tipo $\log I_{i,j}|\Theta$ sono rispettivamente μ_j e v_j (v. (4.9), (4.10)).

Si ha per $\log P_{i,j}|\Theta$, dalle (4.3)

$$\begin{aligned}\mathbf{E}(\log P_{i,j}) &= \sum_{m=0}^j \Phi_m = \eta_j, \\ \mathbf{Var}(\log P_{i,j}) &= \sum_{m=0}^j \sigma_j^2 = w_j.\end{aligned}$$

Dunque il valore atteso di (4.14) è $(\mu_I, \eta_j, \mu_j, \dots, \mu_0)$. La matrice di varianza-covarianza è

$$\Sigma = \begin{bmatrix} v_I^2 & w_j^2 & v_I^2 & v_I^2 & \dots & v_I^2 \\ w_j^2 & w_j^2 & w_j^2 & w_j^2 & \dots & w_j^2 \\ v_I^2 & w_j^2 & v_j^2 & v_j^2 & \dots & v_j^2 \\ v_I^2 & w_j^2 & v_j^2 & v_{j-1}^2 & \dots & v_{j-1}^2 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ v_I^2 & w_j^2 & v_j^2 & v_{j-1}^2 & \dots & v_0^2 \end{bmatrix}.$$

Infatti, si ha $Cov(\log I_{i,h}, \log I_{i,k}|\Theta) = v_h^2$, (v. (4.11)), e dove

$$w_j^2 = \sum_{m=0}^j \sigma_m^2, \quad l \in \{I, j, \dots, 0\}.$$

Per le (4.3), (4.8), sfruttando la proprietà di linearità della covarianza e la prima delle Ipotesi 4.1, riesce

$$\begin{aligned}\mathbf{Cov}(\log P_{i,j}, \log I_{i,l}|\Theta) &= Cov\left(\sum_{m=0}^j \xi_{i,m}, \sum_{k=0}^I \xi_{i,k} + \sum_{h=j}^{I-1} (-\zeta_{i,h})|\Theta\right) = \\ &= \sum_{m=0}^j \sum_{k=0}^I Cov(\xi_{i,m}, \xi_{i,k}|\Theta) - \sum_{m=0}^j \sum_{h=j}^{I-1} Cov(\zeta_{i,m}, \zeta_{i,h}|\Theta) = \\ &= \sum_{m=0}^j \mathbf{Var}(\xi_{i,m}|\Theta) = \sum_{m=0}^j \sigma_m^2\end{aligned}$$

Poiché condizionatamente a Θ grandezze relative a diversi anni di origine sono stocasticamente indipendenti, dell'informazione portata da \mathcal{B}_j è rilevante sulla valutazione probabilistica di $I_{i,I}$ solo l'informazione portata dalle grandezze dell'anno di origine i . Possiamo allora applicare il Lemma 4.1 prendendo

$$\begin{aligned}X_1 &= \log I_{i,I}|\{\Theta, \mathcal{B}_j\}, \\ X^{(2)} &= (\log P_{i,j}, \log I_{i,j}, \dots, \log I_{i,0})'|\{\Theta\},\end{aligned}$$

con parametri

$$\begin{aligned} m_1 &= \mu_I, \\ m^{(2)} &= (\eta_j, \mu_j, \dots, \mu_0), \\ \Sigma_{1,1} &= v_I^2, \\ \Sigma_{1,2} &= (w_j^2, v_I^2, \dots, v_I^2) \in \mathbb{R}^{j+2}, \end{aligned}$$

e $\Sigma_{2,2} = \mathbf{Var}(X^{(2)})$ data da

$$\Sigma_{2,2} = \begin{bmatrix} w_j^2 & w_j^2 & w_j^2 & w_j^2 & \dots & w_j^2 \\ w_j^2 & v_j^2 & v_j^2 & v_j^2 & \dots & v_j^2 \\ w_j^2 & v_j^2 & v_{j-1}^2 & v_{j-1}^2 & \dots & v_{j-1}^2 \\ w_j^2 & v_j^2 & v_{j-1}^2 & v_{j-2}^2 & \dots & v_{j-2}^2 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ w_j^2 & v_j^2 & v_{j-1}^2 & v_{j-2}^2 & \dots & v_0^2 \end{bmatrix}.$$

La struttura della matrice $\Sigma_{2,2}$ è simile a quella trovata nella Proposizione 4.1 e analogo è il calcolo dell'inversa. La matrice inversa $\Sigma_{2,2}^{-1}$ ha la seguente proprietà (v. Merz e Wüthrich (2010))

$$\Sigma_{1,2}\Sigma_{2,2}^{-1} = \left(\frac{v_j^2 - v_I^2}{v_j^2 - w_j^2}, \frac{v_I^2 - w_j^2}{v_j^2 - w_j^2}, 0, \dots, 0 \right) = (1 - \beta_j, \beta_j, 0, \dots, 0) \in \mathbb{R}^{j+2}.$$

Osserviamo che $v_j^2 > v_I^2 > w_j^2$, perciò $0 < \beta_j < 1$. Abbiamo allora che

$$m_1 + \Sigma_{1,2}\Sigma_{2,2}^{-1}(X^{(2)} - m^{(2)}) = \mu_I + (1 - \beta_j)(\log P_{i,j} - \eta_j) + \beta_j(\log I_{i,j} - \mu_j)$$

e

$$\Sigma_{1,1} - \Sigma_{1,2}\Sigma_{2,2}^{-1}\Sigma_{2,1} = v_I^2 - (1 - \beta_j)w_j^2 - \beta_jv_I^2 = (1 - \beta_j)(v_I^2 - w_j^2).$$

Applicando il Lemma 4.1 si ha la tesi. \square

Osservazione. La distribuzione condizionale di $\log P_{i,I} | \{\Theta, \mathcal{B}_j\}$ varia in funzione delle osservazioni $\log P_{i,j}$ e $\log I_{i,j}$. Più precisamente, la speranza matematica a priori μ_I per il costo ultimo $P_{i,I} = I_{i,I}$ viene aggiornata con una media ponderata tra il residuo dei pagamenti ($\log P_{i,j} - \eta_j$) e il residuo degli *incurred losses* ($\log I_{i,j} - \mu_j$), con peso di credibilità

$$\beta_j = \frac{\sum_{m=j+1}^I \sigma_m^2}{\sum_{m=j+1}^I \sigma_m^2 + \sum_{n=j}^{I-1} \tau_n^2}.$$

Analogamente, la varianza a priori $v_I^2 - w_j^2$ viene aggiornata con peso di credibilità $(1 - \beta_j)$. \square

Dal Teorema 4.1 segue il seguente corollario, che permette di ottenere la speranza matematica del costo ultimo usando l'informazione congiunta.

Corollario 4.2 (Costo ultimo PIC). *Sotto le ipotesi del metodo PIC otteniamo*

$$\mathbf{E}(P_{i,I}|\{\mathcal{B}_j, \Theta\}) = P_{i,j} \cdot \exp \left\{ \sum_{l=j+1}^I \left(\Phi_l + \frac{\sigma_l^2}{2} \right) \right\} \cdot \exp \left\{ \beta_j \left(\log \frac{I_{i,j}}{P_{i,j}} - (\mu_j - \eta_j) - \sum_{l=j+1}^I \frac{\sigma_l^2}{2} \right) \right\}, \quad (4.15)$$

o, equivalentemente,

$$\mathbf{E}(P_{i,I}|\{\mathcal{B}_j, \Theta\}) = I_{i,j} \cdot \exp \left\{ \sum_{l=j}^{I-1} \Psi_l \right\} \cdot \exp \left\{ (1 - \beta_j) \left(\log \frac{P_{i,j}}{I_{i,j}} - (\eta_j - \mu_j) + \sum_{l=j+1}^I \frac{\sigma_l^2}{2} \right) \right\}. \quad (4.16)$$

Dimostrazione. Dal Teorema 4.1 abbiamo che, poiché $\log P_{i,j}|\{\Theta, \mathcal{B}_j\}$ ha distribuzione normale, $P_{i,I}|\{\Theta, \mathcal{B}_j\}$ ha distribuzione normale di valore atteso

$$\mathbf{E} \left(P_{i,I}|\{\Theta, \mathcal{B}_j\} \right) = \exp \left\{ \mu_I + (1 - \beta_j)(\log P_{i,j} - \eta_j) + \beta_j(\log I_{i,j} - \mu_j) + (1 - \beta_j) \frac{v_I^2 - w_j^2}{2} \right\} \quad (4.17)$$

Con semplici passaggi, essendo

$$\begin{aligned} \mu_I - \eta_j &= \sum_{m=0}^I \Phi_m - \sum_{m=0}^j \Phi_m = \sum_{m=j+1}^I \Phi_m, \\ v_I^2 - w_j^2 &= \sum_{m=0}^I \sigma_m^2 - \sum_{m=0}^j \sigma_m^2 = \sum_{m=j+1}^I \sigma_m^2, \end{aligned}$$

si ottiene

$$\mathbf{E}(P_{i,I}|\{\Theta, \mathcal{B}_j\}) = P_{i,j} \exp \left\{ \sum_{l=j+1}^I \left(\Phi_l + \frac{\sigma_l^2}{2} \right) \right\} \exp \left\{ \beta_j \left(\eta_j - \log P_{i,j} + \log I_{i,j} - \mu_j - \frac{v_I^2 - w_j^2}{2} \right) \right\},$$

ovvero la (4.15).

Riprendiamo ora la speranza matematica di $P_{i,I}|\{\Theta, \mathcal{B}_j\}$ espresse dalla (4.17) e nell'esponente sommiamo e sottraiamo $\log I_{i,j}$ e μ_j , otteniamo

$$\begin{aligned} \mathbf{E}(P_{i,I}|\{\Theta, \mathcal{B}_j\}) &= \\ &= \exp \left\{ \mu_I - \mu_j + (1 - \beta_j) \log P_{i,j} - (1 - \beta_j) \eta_j - (1 - \beta_j) \log I_{i,j} + \log I_{i,j} \right. \\ &\quad \left. + (1 - \beta_j) \mu_j + (1 - \beta_j) \frac{v_I^2 - w_j^2}{2} \right\} = \\ &= I_{i,j} \exp \left\{ \mu_I - \mu_j + (1 - \beta_j) \left(\log \frac{P_{i,j}}{I_{i,j}} - (\eta_j - \mu_j) + \frac{v_I^2 - w_j^2}{2} \right) \right\} = \\ &= I_{i,j} \exp \left\{ \sum_{l=j}^{I-1} \Psi_l \right\} \exp \left\{ (1 - \beta_j) \left(\log \frac{P_{i,j}}{I_{i,j}} - (\eta_j - \mu_j) + \sum_{l=j+1}^I \frac{\sigma_l^2}{2} \right) \right\} \end{aligned}$$

dove abbiamo sfruttato la

$$\mu_I - \mu_j = \sum_{l=j}^{I-1} \Psi_l.$$

□

Osservazione. Sfruttando l'informazione congiunta otteniamo quindi un termine correttivo

$$\exp \left\{ \beta_j \left(\log \frac{I_{i,j}}{P_{i,j}} - (\mu_j - \eta_j) - \sum_{l=j+1}^I \frac{\sigma_l^2}{2} \right) \right\}$$

rispetto all'espressione ottenuta per $\mathbf{E}(P_{i,I}|\{\Theta, \mathcal{B}_j^P\})$. Tale fattore di aggiustamento è funzione dei residui dei rapporti $(I/P)_{i,j}$: maggiore il residuo del rapporto $(I/P)_{i,j}$ e maggiore sarà il termine correttivo. Questo meccanismo di aggiustamento è simile a quello incontrato nel modello MCL.

4.3 Semplice esemplificazione numerica

Le espressioni (4.15), (4.16) consentirebbero di stimare i costi ultimi, data l'informazione disponibile in I: $\mathcal{D}_I = \{P_{i,j}, I_{i,j}, i + j \leq I\}$, se i parametri Θ fossero noti. In tal caso, lo stimatore bayesiano del costo ultimo $\mathbf{E}(P_{i,I}|\mathcal{D}_I)$ è infatti

$$\tilde{P}_{i,I} = \mathbf{E}(P_{i,I}|\mathcal{D}_I) = \mathbf{E}(P_{i,I}|\{\Theta, \mathcal{B}_{I-i}\}).$$

Consideriamo allora un esempio molto semplice che permetta di comprendere l'algoritmo del metodo supponendo Θ noto. Consideriamo i seguenti triangoli di run-off con $I = 1$:

Tabella 4.1: triangoli di run-off.

	$P_{i,j}$		$I_{i,j}$		
	0	1	0	1	
0	22,026	24,343	0	25,002	23,530
1	21,789		1	22,805	

Sia Θ noto e pari a $\Theta = (\Phi_0, \Phi_1, \Psi_0, \sigma_0, \sigma_1, \tau_0) = (10.02, 0.03, -0.02, 0.14, 0.08, 0.06)$. Il nostro scopo è la stima del costo ultimo data l'informazione congiunta, ovvero l'applicazione della formula (4.15). Procediamo allora al calcolo delle grandezze coinvolte:

$$\mu_0 = \Phi_0 + \Phi_1 - \Psi_0 = 10.07$$

$$\mu_1 = \Phi_0 + \Phi_1 = 10.05$$

$$v_0^2 = \sigma_0^2 + \sigma_1^2 + \tau_0^2 = 0.0296$$

$$v_1^2 = \sigma_0^2 + \sigma_1^2 = 0.026$$

$$\begin{aligned}\eta_0 &= \Phi_0 = 10.02 \\ \eta_1 &= \Phi_0 + \Phi_1 = 10.05 \\ w_0^2 &= \sigma_0^2 = 0.0196 \\ w_1^2 &= \sigma_0^2 + \sigma_1^2 = 0.026 \\ \beta_0 &= \frac{v_1^2 - w_0^2}{v_0^2 - w_0^2} = 0.64\end{aligned}$$

La previsione del costo ultimo sarà allora:

$$\begin{aligned}\mathbf{E}(P_{1,1}|\{\mathcal{B}_0, \Theta\}) &= \mathbf{E}(I_{1,1}|\{\mathcal{B}_0, \Theta\}) = \\ &= P_{1,0} \cdot \exp\left\{\Phi_1 + \frac{\sigma_1^2}{2}\right\} \cdot \exp\left\{\beta_0 \left(\log \frac{I_{1,0}}{P_{1,0}} - (\mu_0 - \eta_0) - \frac{\sigma_1^2}{2}\right)\right\} = \\ &= 22,414.88.\end{aligned}$$

Possiamo anche calcolare i costi ultimi dati dalle equazioni (4.5) e (4.12), i quali utilizzano un solo canale di informazione. Utilizzando solo l'informazione sui pagamenti abbiamo

$$\mathbf{E}(P_{1,1}|\{\mathcal{B}_0^P, \Theta\}) = P_{1,0} \cdot \exp\left\{\Phi_1 + \frac{\sigma_1^2}{2}\right\} = 22,524.54,$$

mentre utilizzando solo informazione sugli *incurred losses* si ha

$$\begin{aligned}\mathbf{E}(P_{1,1}|\{\mathcal{B}_0^I, \Theta\}) &= I_{1,0} \cdot \exp\left\{\Psi_0 + \frac{\tau_0^2}{2}\right\} \cdot \exp\left\{\alpha_0 \left(\mu_0 - \log I_{1,0} - \frac{\tau_0^2}{2}\right)\right\} = 22,348.72, \\ \text{con } \alpha_0 &= 1 - \frac{v_1^2}{v_0^2} = 0.1216.\end{aligned}$$

Osserviamo che $\mathbf{E}(P_{1,1}|\{\mathcal{B}_0, \Theta\})$ è compreso tra $\mathbf{E}(P_{1,1}|\{\mathcal{B}_0^P, \Theta\})$ e $\mathbf{E}(P_{1,1}|\{\mathcal{B}_0^I, \Theta\})$.

4.4 Stima dei parametri

Nella pratica i parametri Θ non sono noti e devono essere stimati dai dati. Per farlo è usato un approccio Bayesiano: si assume una distribuzione a priori per i parametri e si calcola la distribuzione a posteriori date le osservazioni. Siano

$$\begin{aligned}\mathcal{D}_I^P &= \{P_{i,j} : i + j \leq I\}, \\ \mathcal{D}_I^I &= \{I_{i,j} : i + j \leq I\}, \\ \mathcal{D}_I &= \mathcal{D}_I^P \cup \mathcal{D}_I^I,\end{aligned}$$

le osservazioni disponibili al tempo I .

Ipotesi 4.2 (Bayesian PIC). Si assume che le Ipotesi 4.1 siano verificate con $\sigma_0, \dots, \sigma_I$ e $\tau_0, \dots, \tau_{I-1}$ deterministici e che riesca

$$\begin{aligned}\Phi_m &\sim N(\phi_m, s_m^2), & m = 0, \dots, I, \\ \Psi_n &\sim N(\psi_n, t_n^2), & n = 0, \dots, I-1.\end{aligned}$$

□

Si suppone dunque che i parametri $\sigma_0, \dots, \sigma_I$ e $\tau_0, \dots, \tau_{I-1}$ siano deterministici e noti, anziché supporre una distribuzione a priori per l'intero vettore Θ . Questo ha il vantaggio di permettere il calcolo di forme chiuse per le distribuzioni a posteriori. Svolgeremo ora il caso in cui si utilizzi l'informazione congiunta \mathcal{D}_I ; in Merz e Wüthrich (2010) vengono svolti anche i casi in cui si utilizzi solo \mathcal{D}_I^P o solo \mathcal{D}_I^I . Per le dimostrazioni in questo paragrafo si rimanda all'articolo Merz e Wüthrich (2010).

La distribuzione a posteriori $u(\Theta|\mathcal{D}_I)$ di Θ dato \mathcal{D}_I è la seguente:

$$u(\Theta|\mathcal{D}_I) \propto l_{\mathcal{D}_I}(\Theta) \cdot \prod_{m=0}^I \exp\left\{-\frac{1}{2s_m^2}(\Phi_m - \phi_m)^2\right\} \cdot \prod_{n=0}^{I-1} \exp\left\{-\frac{1}{2t_n^2}(\Psi_n - \psi_n)^2\right\},$$

con $l_{\mathcal{D}_I}(\Theta)$ data da

$$\begin{aligned}l_{\mathcal{D}_I}(\Theta) &= \prod_{j=0}^I \prod_{i=0}^{I-j} \frac{1}{\sqrt{2\pi}\sigma_j P_{i,j}} \exp\left\{-\frac{1}{2\sigma_j^2}\left(\Phi_j - \log \frac{P_{i,j}}{P_{i,j-1}}\right)^2\right\} \\ &\cdot \prod_{i=1}^I \frac{1}{\sqrt{2\pi}(v_{I-j}^2 - w_{I-i}^2)I_{i,I-i}} \\ &\cdot \exp\left\{-\frac{1}{2(v_{I-i}^2 - w_{I-i}^2)}\left(\mu_{I-i} - \eta_{I-i} - \log \frac{I_{i,I-i}}{P_{i,I-i}}\right)^2\right\} \\ &\cdot \prod_{j=0}^{I-1} \prod_{i=0}^{I-j-1} \frac{1}{\sqrt{2\pi}\tau_j I_{i,j}} \exp\left\{-\frac{1}{2\tau_j^2}\left(\Psi_j + \log \frac{I_{i,j}}{I_{i,j+1}}\right)^2\right\}.\end{aligned}$$

Abbiamo allora il seguente risultato:

Teorema 4.2. *Sotto le Ipotesi 4.2 la distribuzione a posteriori $u(\Theta|\mathcal{D}_I)$ è una Normale multivariata di media a posteriori $\theta^{post}(\mathcal{D}_I)$ e matrice di varianza-covarianza a posteriori $\Sigma(\mathcal{D}_I)$. La matrice inversa $\Sigma(\mathcal{D}_I)^{-1} = (a_{m,n})_{0 \leq n, m \leq 2I}$ è data da*

$$\begin{aligned}a_{n,m} &= (s_n^{-2} + (I - n + 1)\sigma_n^{-2})1_{\{n=m\}} + \sum_{i=0}^{\min(n-1, m-1)} (v_i^2 - w_i^2)^{-1}, & n, m = 0, \dots, I, \\ a_{I+1+n, I+1+m} &= (t_n^{-2} + (I - n)\tau_n^{-2})1_{\{n=m\}} + \sum_{i=0}^{\min(n, m)} (v_i^2 - w_i^2)^{-1}, & n, m = 0, \dots, I-1,\end{aligned}$$

$$a_{n,I+1+m} = a_{I+1+m,n} = - \sum_{i=0}^{\min(n-1,m)} (v_i^2 - w_i^2)^{-1}, \quad n = 0, \dots, I, \quad m = 0, \dots, I-1.$$

La media a posteriori $\theta^{post}(\mathcal{D}_I) = (\phi_0^{post}, \dots, \phi_I^{post}, \psi_0^{post}, \dots, \psi_{I-1}^{post})'$ è data da

$$\theta^{post}(\mathcal{D}_I) = \Sigma(\mathcal{D}_I)(c_0, \dots, c_I, b_0, \dots, b_{I-1})',$$

con il vettore $(c_0, \dots, c_I, b_0, \dots, b_{I-1})$ dato da

$$c_j = s_j^{-2} \phi_j + \sigma_j^{-2} \sum_{i=0}^{I-j} \log \frac{P_{i,j}}{P_{i,j-1}} + \sum_{i=I-j+1}^I (v_{I-i}^2 - w_{I-i}^2)^{-1} \log \frac{I_{i,I-i}}{P_{i,I-i}},$$

$$b_j = t_j^{-2} \psi_j - \tau_j^{-2} \sum_{i=0}^{I-j-1} \log \frac{I_{i,j}}{I_{i,j+1}} - \sum_{i=I-j}^I (v_{I-i}^2 - w_{I-i}^2)^{-1} \log \frac{I_{i,I-i}}{P_{i,I-i}}.$$

Dal Teorema 4.2 discende il seguente corollario, che permette il calcolo dei costi ultimi dato \mathcal{D}_I :

Corollario 4.3. *Sotto le Ipotesi 4.2 si ha*

$$\mathbf{E}(P_{i,I} | \mathcal{D}_I) = P_{i,I-i}^{1-\beta_{I-i}} \cdot I_{i,I-i}^{\beta_{I-i}} \cdot \exp \left\{ (1 - \beta_{I-i}) \sum_{l=I-i+1}^I \phi_l^{post} + \beta_{I-i} \sum_{l=I-i}^{I-1} \psi_l^{post} \right\} \cdot \exp \left\{ (1 - \beta_{I-i}) \frac{v_I^2 - w_{I-i}^2}{2} + \frac{(s_i^{post})^2}{2} \right\}, \quad (4.18)$$

dove $(s_i^{post})^2 = \mathbf{e}_i' \Sigma(\mathcal{D}_I) \mathbf{e}_i$, con

$$\mathbf{e}_i = (0, \dots, 0, 1 - \beta_{I-i}, \dots, 1 - \beta_{I-i}, 0, \dots, 0, \beta_{I-i}, \dots, \beta_{I-i})' \in \mathbb{R}^{2I+1}.$$

Rimangono da stimare i parametri deterministici σ_j^2 e τ_j^2 ; per essi si possono usare i seguenti stimatori:

$$\hat{\sigma}_j^2 = \frac{1}{I-j} \sum_{i=0}^{I-j} \left(\log \frac{P_{i,j}}{P_{i,j-1}} - \hat{\Phi}_j \right)^2, \quad j = 0, \dots, I-1,$$

con $\hat{\Phi}_j = \frac{1}{I-j+1} \sum_{i=0}^{I-j} \log \frac{P_{i,j}}{P_{i,j-1}}$ e

$$\hat{\tau}_j^2 = \frac{1}{I-j-1} \sum_{i=0}^{I-j-1} \left(\log \frac{I_{i,j}}{I_{i,j+1}} - \hat{\Psi}_j \right)^2, \quad j = 0, \dots, I-1,$$

con $\hat{\Psi}_j = \frac{1}{I-j} \sum_{i=0}^{I-j-1} \log \frac{I_{i,j}}{I_{i,j+1}}$. $\hat{\sigma}_I$ e $\hat{\tau}_I$ è possibile stimarli ad esempio con una regressione log-lineare. La speranza matematica a posteriori $\mathbf{E}(P_{i,I} | \mathcal{D}_I)$ può essere ottenuta per stimare i costi ultimi, si assume dunque come stimatore di tali costi lo stimatore bayesiano $\hat{P}_{i,I} = \mathbf{E}(P_{i,I} | \mathcal{D}_I)$.

Valutazione dell'incertezza

Sotto le Ipotesi 4.2, il metodo PIC permette di ricavare in forma analitica il msep condizionato.

Teorema 4.3. *Sotto le Ipotesi 4.2, data l'informazione \mathcal{D}_I , si ha che*

$$\begin{aligned} \text{msep}_{\sum_{i=1}^I P_{i,I}|\mathcal{D}_I} \left(\mathbf{E} \left[\sum_{i=1}^I P_{i,I} \middle| \mathcal{D}_I \right] \right) &= \\ &= \sum_{1 \leq i, k \leq I} \left(\exp \left((1 - \beta_{I-i})(v_I^2 - w_{I-i}^2) 1_{\{i=k\}} + (\mathbf{e}_i)' \Sigma(\mathcal{D}_I) \mathbf{e}_k \right) - 1 \right) \\ &\cdot \mathbf{E}(P_{i,I}|\mathcal{D}_I) \cdot \mathbf{E}(P_{k,I}|\mathcal{D}_I), \end{aligned}$$

con $\mathbf{E}(P_{i,I}|\mathcal{D}_I)$ data dalla (4.18).

4.5 Esempificazione numerica

In questa applicazione ricalcheremo l'esempio in Merz e Wüthrich (2010), applicandolo a un nuovo set di dati, precisamente ai dati in Wüthrich e Merz (2008) già utilizzati nel Capitolo 3. Consideriamo distribuzioni a priori non informative per le variabili aleatorie Φ_j e Ψ_j , supponendo quindi che $s_m^2 \rightarrow \infty$ e che $t_m^2 \rightarrow \infty$. In questo modo non abbiamo bisogno di alcuna conoscenza a priori, in quanto i valori ϕ_m e ψ_n scompaiono dalle formule che definiscono c_m e b_n . Per applicare il metodo servirà solo stimare i parametri deterministici σ_j e τ_j .

Otteniamo le stime per σ_j e τ_j nella Tabella 4.2, dove i valori in parentesi sono stati ricavati per mezzo di una regressione log-lineare. Nella Tabella 4.3 sono riportati i risultati per i costi ultimi e per le riserve per anno di origine, mentre nella Tabella 4.4 è riportato il risultato per la riserva di portafoglio. Quest'ultima è pari a 53,638, valore confrontabile con i risultati ottenuti per questo dataset con il metodo MCL, pari a 52,034 (*paid*) e 54,628 (*incurred*).

Come in precedenza evidenziato, sotto le Ipotesi 4.2 il metodo PIC ha il vantaggio di permettere il calcolo in forma chiusa del msep condizionato, che risulta pari a 5,044.45.

Tabella 4.2: stime di σ_j e τ_j .

j	σ_j	τ_j
0	0.1501	0.0319
1	0.0331	0.0181
2	0.0173	0.0118
3	0.0081	0.0128
4	0.0047	0.0106
5	0.0045	0.0074
6	0.0047	0.0046
7	0.0018	0.0023
8	0.0052	0.0012
9	0.0024	0.0014
10	0.0003	0.0016
11	0.0004	0.0103
12	0.0018	0.0003
13	0.0002	0.0004
14	0.0001	0.0001
15	0.0002	(0.0002)
16	(0.0001)	

Tabella 4.3: stime del costo ultimo e delle riserve per anno di origine.

i	$\mathbf{E}(P_{i,I} \mathcal{D}_I)$	\hat{R}_i
1	74,905	7
2	86,265	-21
3	95,723	48
4	93,899	88
5	94,064	216
6	93,899	197
7	90,049	217
8	93,152	322
9	124,122	676
10	113,744	764
11	114,779	974
12	104,267	1,248
13	113,240	2,340
14	103,361	2,914
15	109,020	5,968
16	110,304	37,680

Tabella 4.4: stima delle riserva di portafoglio.

\hat{R}
53,638

Capitolo 5

Modello individuale per Paid and Incurred losses

In questo capitolo presentiamo il modello di valutazione della riserva sinistri proposto da Pigeon et al. (2014). Si tratta di un modello che tiene conto delle informazioni sia sui pagamenti sia sugli *incurred losses* ed estende un modello di Pigeon et al. (2013), introdotto solo per i pagamenti. Rispetto ai modelli descritti nei capitoli precedenti che trattano dati aggregati riassunti in triangoli di run-off, l'attuale modello è di tipo individuale, nel senso che le informazioni che si utilizzano sono relative ai singoli sinistri di un portafoglio. Richiede di disporre di dati piuttosto dettagliati che non ci è stato possibile reperire, ciò ha impedito di sviluppare una applicazione numerica del modello.

Può essere visto come una estensione individuale del modello di Merz e Wüthrich (2010) descritto nel Capitolo 4.

5.1 Dati e notazioni

Il modello ipotizza vi sia a disposizione un dataset sufficientemente dettagliato riguardo allo sviluppo dei pagamenti e degli *incurred losses* per ciascun sinistro. In particolare occorre conoscere:

- data di accadimento del sinistro;
- data di denuncia del sinistro;
- date e importi dei pagamenti effettuati;
- date e importi degli *incurred losses* (dei quali vi è generalmente una stima iniziale e successive modifiche);
- data di chiusura del sinistro.

Il modello è sviluppato a tempo discreto, con periodi di un anno. Sia k il k -esimo sinistro nel dataset e I il numero di periodi di accadimento considerati. Siano $\{P_{k,1}, P_{k,2}, \dots\}$ e

$\{I_{k,1}, I_{k,2}, \dots\}$ i processi stocastici a tempo discreto per i pagamenti e gli *incurred losses* del sinistro k -esimo. Indichiamo con:

- K_i il numero di sinistri nel periodo i , $i = 1, \dots, I^1$.
- $\omega(i)$ una misura di esposizione per il periodo i , $i = 1, \dots, I$.
- T_k ritardo della denuncia del sinistro k , ovvero il numero di periodi tra il periodo di accadimento e il periodo di notifica alla compagnia; $T_k = 0$ se il sinistro viene denunciato nello stesso periodo di accadimento.

Per il processo dei pagamenti del sinistro k introduciamo le seguenti variabili aleatorie:

- Q_k^P il ritardo del primo pagamento, ovvero il numero di periodi tra la denuncia e il primo pagamento $\neq 0$.
- U_k^P il numero di periodi con pagamenti incrementali $\neq 0$, dopo il primo periodo con pagamento positivo.
- $P_{k,1}, P_{k,2}, \dots, P_{k,U_k^P+1}$ i pagamenti cumulati per il sinistro k ; $P_{k,j} \neq P_{k,j-1}$ poiché consideriamo solo pagamenti incrementali $\neq 0$; si noti che in $P_{k,j}$, l'indice j si riferisce al j -esimo pagamento cumulato, non al differimento j -esimo come è usuale;
- $N_{k,j}^P$ il ritardo tra due periodi con pagamento, ovvero il numero di periodi tra il pagamento j e $j+1$, $j = 1, \dots, U_k^P+1$; $N_{k,U_k^P+1}^P$ è il numero di periodi tra l'ultimo pagamento e la chiusura del sinistro.

Analogamente, per il processo degli *incurred losses* introduciamo:

- Q_k^I il ritardo della stima dell'inventario iniziale, ovvero il numero di periodi tra la denuncia del sinistro e la stima iniziale; normalmente $Q_k^I = 0$, ma per motivi legislativi o altri fattori esterni la stima iniziale potrebbe richiedere più tempo.
- U_k^I il numero di periodi con modifiche degli *incurred losses*, dopo il periodo nel quale è stata effettuata la prima valutazione.
- $I_{k,1}, I_{k,2}, \dots, I_{k,U_k^I+1}$ gli *incurred losses* per il sinistro k . Anche in questo caso consideriamo $I_{k,j} \neq I_{k,j-1}$.
- $N_{k,j}^I$ il ritardo tra due periodi con modifica degli *incurred losses*, $j = 1, \dots, U_k^I+1$; $N_{k,U_k^I+1}^I$ è il numero di periodi tra l'ultima modifica e la chiusura del sinistro.

Sia τ l'istante di valutazione delle riserve. Indichiamo con \mathcal{I}_τ^P e \mathcal{I}_τ^I l'informazione sui pagamenti e, rispettivamente, sugli *incurred losses* disponibile in τ . Poniamo $\mathcal{I}_\tau := \mathcal{I}_\tau^P \cup \mathcal{I}_\tau^I$ l'informazione congiunta disponibile in τ .

¹Segnaliamo che in questo capitolo gli anni di origine sono codificati a partire da 1, anziché da 0.

Esempio. Nell'esempio numerico illustrato nella Tabella 5.1 viene mostrato come a partire dai dati sia possibile ricavare, per un fissato sinistro, le determinazioni dei numeri che descrivono i processi dei pagamenti e degli *incurred losses*. I dati sono riportati nella prima parte della tabella: P indica l'importo dei pagamenti cumulati, I gli *incurred losses*, aggregati per anno di pagamento. Nell'ultima parte sono riportate le corrispondenti determinazioni delle grandezze sopra definite. Oltre ai valori in tabella, riesce: $T = 0$, $Q^P = 1$, $U^P = 2$, $N_1^P = 1$, $N_2^P = 1$, $N_3^P = 1$, $Q^I = 0$, $U^I = 3$, $N_1^I = 2$, $N_2^I = 1$, $N_3^I = 1$, $N_4^I = 0$. Notiamo che i dati sono raccolti in tempo continuo, mentre il modello è a tempo discreto.

Tabella 5.1: dati disponibili per un singolo sinistro.

Dati	Determinazioni	
Data di accadimento	1/5/2009	
Data di denuncia	26/9/2009	
$P = 0, I = 7,300$	12/12/2009	$I_1 = 7,300$
$P = 1,000$	6/3/2010	
$P = 1,600$	25/4/2010	$P_1 = 1,600$
$P = 3,350, I = 7,800$	7/1/2011	$P_2 = 3,350, I_2 = 7,800$
$P = 6,643, I = 6,900$	5/5/2012	$P_3 = 6,643, I_3 = 6,900$
$P = 6,643, I = 6,643$	10/7/2013	$I_4 = 6,643$
Data di chiusura	10/7/2013	

□

5.2 Il modello

5.2.1 Sviluppo dei pagamenti

Dato un sinistro k e $U_k^P = u_k^P$ (dunque condizionatamente a $U_k^P = u_k^P$), ipotizziamo che i pagamenti cumulati $P_{k,j}$ siano legati dalla seguente formula ricorsiva:

$$P_{k,j} = P_{k,j-1} \cdot \lambda_{j-1}^{(k)}, \quad j = 2, \dots, u_k^P + 1, \quad (5.1)$$

con valore iniziale $P_{k,1}$. I $\lambda_j^{(k)}$ sono *link ratios tra pagamenti*. A differenza dei fattori di sviluppo del metodo *chain-ladder* sono numeri aleatori, inoltre l'indice j si riferisce al pagamento j -esimo anziché al periodo di differimento j -esimo. L'informazione sui pagamenti può essere allora riassunta nel vettore

$$\left(P_{k,1} \quad \lambda_1^{(k)} \quad \lambda_2^{(k)} \quad \dots \quad \lambda_{u_k^P}^{(k)} \right)'. \quad (5.2)$$

Il pagamento $P_{k,j}$ può essere ottenuto con

$$P_{k,j} = P_{k,1} \cdot \prod_{i=1}^{j-1} \lambda_i^{(k)}. \quad (5.3)$$

Collegiamo ora il processo dei pagamenti con il processo degli *incurred losses*, imponendo, condizionatamente a $U_k^P = u_k^P$ e $U_k^I = u_k^I$, la seguente condizione:

$$I_{k,u_k^I+1} = P_{k,u_k^P+1}. \quad (5.4)$$

Tale condizione garantisce che, per il fissato sinistro, il pagamento finale coincida con l'ultimo *incurred loss*. Notiamo che volendo considerare congiuntamente i processi dei pagamenti e degli *incurred losses*, condizionatamente all'informazione sul numero di periodi di pagamento e sul numero di modifiche, rispettivamente, dobbiamo formulare le ipotesi sullo sviluppo dei $P_{k,j}$ e $I_{k,j}$ condizionatamente alla coppia U_k^P, U_k^I . Dunque le (5.1), (5.3) e le relazioni che seguiranno sono da intendersi condizionate a $(U_k^P = u_k^P, U_k^I = u_k^I)$. Per il processo degli *incurred losses* ipotizziamo una ricorsione all'indietro, ovvero

$$I_{k,j-1} = I_{k,j} \cdot \gamma_{j-1}^{(k)}, \quad j = 2, \dots, u_k^I + 1, \quad (5.5)$$

con valore iniziale I_{k,u_k^I+1} ottenibile da (5.4). I $\gamma_j^{(k)}$ sono *link ratios*, aleatori, tra successive modifiche degli *incurred losses*.

Pertanto si ha

$$\begin{aligned} I_{k,j} &= I_{k,u_k^I+1} \cdot \prod_{h=j}^{u_k^I} \gamma_h^{(k)} \\ &= P_{k,u_k^P+1} \cdot \prod_{h=j}^{u_k^I} \gamma_h^{(k)} \\ &= P_{k,1} \cdot \prod_{i=1}^{u_k^P} \lambda_i^{(k)} \prod_{h=j}^{u_k^I} \gamma_h^{(k)}. \end{aligned} \quad (5.6)$$

Il processo per il sinistro k , condizionatamente a $(U_k^P = u_k^P, U_k^I = u_k^I)$, è allora descrivibile dal seguente vettore:

$$\left(P_{k,1} \quad \lambda_1^{(k)} \quad \lambda_2^{(k)} \quad \dots \quad \lambda_{u_k^P}^{(k)} \quad \gamma_{u_k^I}^{(k)} \quad \dots \quad \gamma_1^{(k)} \right)'. \quad (5.7)$$

Le condizioni precedenti devono valere per ogni u_k^P, u_k^I determinazioni possibili di U_k^P, U_k^I e per ogni k .

Esempio. In riferimento al sinistro dell'esempio illustrato nella Tabella 5.1, i valori osservati degli elementi di tale vettore sono (i fattori di sviluppo sono arrotondati alla seconda cifra decimale)

$$P_{k,1} = 1,600 \quad \lambda_1^{(k)} = 2.09 \quad \lambda_2^{(k)} = 1.98 \quad \gamma_3^{(k)} = 1.04 \quad \gamma_2^{(k)} = 1.13 \quad \gamma_1^{(k)} = 0.94.$$

□

5.2.2 Ipotesi probabilistiche

La dinamica temporale dello sviluppo del sinistro k viene modellizzata attraverso le variabili aleatorie $T_k, U_k^P, Q_k^P, N_k^P, U_k^I, Q_k^I, N_k^I$. Per ciascuna di esse si ipotizza una forma distributiva discreta con componenti degeneri. Per esempio, il *ritardo della denuncia* T_k ha funzione di probabilità

$$f_1(t; \boldsymbol{\nu}) = \sum_{s=0}^p \nu_s \cdot I_s(t) + \left(1 - \sum_{s=0}^p \nu_s\right) \cdot f_{T|T>p}(t) \quad (5.8)$$

dove $I_s(t) = 1$ se $t = s$, 0 altrimenti e $f(t)$ è la funzione di probabilità di una distribuzione discreta di parametri $\nu_{s+1}, \dots, \nu_{s+q}$. Ad esempio, se $p = 1$ la funzione di probabilità di T_k , $f_1(t; \boldsymbol{\nu})$, è

$$\begin{aligned} P(T_k = t) &= \nu_0 I_0(t) + \nu_1 I_1(t) + (1 - \nu_0 - \nu_1) f_{T|T>1}(t) \\ &= \begin{cases} \nu_0 & t = 0 \\ \nu_1 & t = 1, \\ (1 - \nu_0 - \nu_1) f_{T|T>1}(t) & t = 2 \end{cases} \end{aligned}$$

dove, per $t \geq 2$, se la distribuzione condizionata deriva da una distribuzione di Poisson di parametro λ ,

$$f_{T|T>1}(t) = \frac{e^{-\lambda} \lambda^t / t!}{1 - e^{-\lambda} - e^{-\lambda} \lambda}.$$

In tal caso la distribuzione di T_k dipende dai parametri (ν_0, ν_1, λ) .

Al fine di specificare il modello, occorre investigare una opportuna distribuzione, ad esempio Poisson, Geometrica, Binomiale Negativa, e un opportuno numero p di componenti degeneri.

Per il numero di sinistri nel periodo i , K_i , si assume una distribuzione di Poisson di parametro $\theta \cdot w(i)$, dove $w(i)$ è una misura di esposizione relativa al periodo i . Occorre tuttavia tenere conto che osserviamo solo i sinistri denunciati. Indichiamo, allora, con D_i il numero di sinistri del periodo di origine i , denunciati entro la data di valutazione della riserva. Posto r_i^* il numero intero di periodi tra il periodo di accadimento e il periodo di valutazione τ e supposto che la valutazione sia effettuata all'inizio dell'anno τ (v. Figura 5.1), il numero aleatorio D_i può essere descritto come segue

$$D_i = \sum_{k=1}^{K_i} |T_k \leq r_i^* - 1|.$$

Avendo supposto che K_i abbia distribuzione di Poisson di parametro $\theta \cdot \omega(i)$, a D_i si assegna distribuzione di Poisson con parametro

$$\theta \cdot w(i) \cdot F_1(r_i^* - 1; \boldsymbol{\nu}),$$

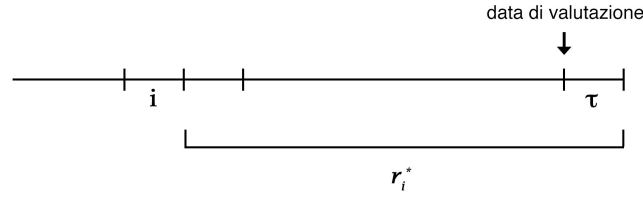


Figura 5.1

dove $F_1(r_i^* - 1; \nu) = P(T_k \leq r_i^* - 1)$. In questo modo teniamo conto della probabilità che il ritardo della denuncia non sia superiore all'intervallo di tempo tra l'accadimento del sinistro e la data di valutazione².

Occorre ora ipotizzare una forma distributiva per il vettore (5.7). L'ipotesi di indipendenza tra *link ratios* è problematica. Questo fornisce la motivazione per prendere in considerazione una distribuzione multidimensionale. La scelta degli Autori nell'articolo citato è di utilizzare una distribuzione detta Multivariate Skew Normal (MSN) che è qui di seguito definita.

Definizione 5.1 (Distribuzione MSN). Sia $\boldsymbol{\mu} = [\mu_1 \dots \mu_l]'$ un vettore di parametri di posizione, $\boldsymbol{\Sigma}$ una matrice di scala ($l \times l$), simmetrica e definita positiva, e $\boldsymbol{\Delta} = [\Delta_1 \dots \Delta_l]'$ un vettore di parametri di forma. Il vettore aleatorio \mathbf{X} ($l \times 1$) segue una distribuzione Multivariate Skew Normal di dimensione l (MSN_l) se la sua funzione di densità è della forma

$$MSN_l(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}^{1/2}, \boldsymbol{\Delta}) = \frac{2^l}{\det(\boldsymbol{\Sigma})^{1/2}} \cdot \phi_l(\boldsymbol{\Sigma}^{-1/2}(\mathbf{x} - \boldsymbol{\mu})) \cdot \prod_{j=1}^l \Phi(\Delta_j \mathbf{e}_j' \boldsymbol{\Sigma}^{-1/2}(\mathbf{x} - \boldsymbol{\mu})) \quad (5.9)$$

dove \mathbf{e}_j è il j -esimo vettore della base canonica di \mathbb{R}^l e

$$\phi_l(y_1, \dots, y_l) = \prod_{j=1}^l \phi(y_j)$$

con $\phi(\cdot)$ e $\Phi(\cdot)$ la funzione di densità e la funzione di ripartizione della Normale standard (unidimensionale).

²Ciò è coerente con l'ipotesi che K_i abbia distribuzione di Poisson di parametro $\theta \cdot \omega(i)$ e che a D_i si assegni distribuzione Poisson-composto. Ricordiamo infatti che se $D = \sum_{k=1}^K |E_k|$ ha distribuzione Poisson-composto (λ, p) , allora D ha distribuzione di Poisson di parametro $\lambda \cdot p$.

Osservazione. La formula (5.9) è valida qualsiasi sia la radice quadrata di Σ , cioè qualunque sia $\Sigma^{1/2}$ tale che $(\Sigma^{1/2})(\Sigma^{1/2})' = \Sigma$. Nella formula (5.12), $\Sigma_v^{1/2}$ è la radice quadrata della matrice Σ_v ottenuta mediante la decomposizione di Cholesky³.

La distribuzione MSN permette di tenere in considerazione la dipendenza tra variabili aleatorie e l'asimmetria, pur rimanendo matematicamente trattabile. Riportiamo un risultato relativo alla funzione generatrice dei momenti delle distribuzioni MSN che sarà utile per alcune verifiche effettuate nel prossimo capitolo.

Lemma 5.1. *La funzione generatrice dei momenti di un vettore aleatorio $\mathbf{X} \sim MSN_l(\boldsymbol{\mu}, \Sigma^{1/2}, \boldsymbol{\Delta})$ in $\mathbf{t} \in \mathbb{R}^l$ è*

$$m_{\mathbf{X}}(\mathbf{t}) = 2^l \cdot \exp\left(\mathbf{t}'\boldsymbol{\mu} + 0.5\mathbf{t}'(\Sigma^{1/2})(\Sigma^{1/2})'\mathbf{t}\right) \cdot \prod_{j=1}^l \Phi\left(\frac{\Delta_j((\Sigma^{1/2})'\mathbf{t})_j}{\sqrt{1 + \Delta_j^2}}\right). \quad (5.10)$$

Torniamo ora al problema di assegnare una valutazione probabilistica ai processi dei pagamenti e degli *incurred losses*. Poniamo $M = \max_k(u_k^P) + \max_k(u_k^I) + 1$, dove $\max_k(u_k^P)$ è la massima determinazione osservata dei numeri aleatori u_k^P , per i sinistri del data set in esame, assunta come massima determinazione possibile del numero di periodi con pagamenti incrementali non nulli, dopo il primo; analogamente per $\max_k(u_k^I)$. Siano inoltre $\boldsymbol{\mu} = [\mu_1 \dots \mu_M]'$ un vettore di posizione,

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1M} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{M1} & \sigma_{M2} & \dots & \sigma_{MM} \end{bmatrix}$$

una matrice di scala e $\boldsymbol{\Delta} = [\Delta_1 \dots \Delta_M]'$ un vettore di parametro di forma. Poniamo

$$\boldsymbol{\Omega} := \left[\log P_1 \quad \log \lambda_1 \quad \log \lambda_2 \quad \dots \quad \log \lambda_{\max_k(u_k^P)} \quad \log \gamma_{\max_k(u_k^I)} \quad \dots \quad \log \gamma_1 \right]'. \quad (5.11)$$

Preso un insieme ordinato $v \subset \{1, 2, \dots, \max_k(u_k^P) + 1, \max_k(u_k^P) + 2, \dots, M\}$, con $|v| \leq M$ dove $|v|$ indica la cardinalità dell'insieme, costruiamo il vettore aleatorio $\boldsymbol{\Omega}_v$ selezionando da $\boldsymbol{\Omega}$ gli elementi corrispondenti agli indici dell'insieme v . Analogamente costruiamo $\boldsymbol{\mu}_v$, $\boldsymbol{\Delta}_v$ e la matrice Σ_v . Assumiamo che

$$\boldsymbol{\Omega}_v \sim MSN_{|v|}(\boldsymbol{\mu}_v, \Sigma_v^{1/2}, \boldsymbol{\Delta}_v). \quad (5.12)$$

Infine, al processo (5.7) condizionato a $(U_k^P = u_k^P, U_k^I = u_k^I)$ assegniamo la legge che deriva dall'assegnare a

$$(\log P_{k,1}, \log \lambda_1^{(k)}, \dots, \log \lambda_{u_k^P}^{(k)}, \log \gamma_{u_k^I}^{(k)}, \dots, \log \gamma_1^{(k)}) \quad (5.13)$$

³Per una matrice Σ simmetrica e definita positiva, la decomposizione di Cholesky è $\mathbf{L}\mathbf{L}' = \Sigma$, con \mathbf{L} matrice triangolare inferiore. Spesso è comodo allora porre $\Sigma^{1/2} = \mathbf{L}$.

la legge di Ω_v , essendo v l'insieme che porta a selezionare nella (5.11) gli elementi corrispondenti agli elementi in (5.13):

$$v = \{1, 2, \dots, u_k^P + 1, \max_k(u_k^P) + 1 + (\max_k(u_k^I) - u_k^I), \dots, M\}, \quad |v| = u_k^P + u_k^I + 1.$$

5.3 Calcolo delle riserve

Le compagnie di assicurazione distinguono tra riserve RBNS (Reported But Not Settled) e IBNR (Incurred But Not Reported). Nel primo caso viene calcolata una riserva per sinistri già denunciati alla compagnia di assicurazione, ma non ancora chiusi. Nel secondo si calcola una riserva per sinistri già accaduti, ma non ancora denunciati alla compagnia.

In questo capitolo distingueremo ulteriormente tra sinistri denunciati ma senza ancora nessun pagamento e sinistri denunciati che hanno già dato luogo a pagamenti. Nel primo caso parliamo di sinistri RBNP (Reported But Not Paid), nel secondo caso di sinistri RBNS puri, o semplicemente RBNS, con rispettive riserve.

Nelle proposizioni seguenti considereremo un generico sinistro e ometteremo l'indice k .

Proposizione 5.1. *Dati $U^P = u^P$ e $U^I = u^I$ e $v = \{1, 2, \dots, u^P, u^P + 1, \dots, u^P + u^I + 1\}$, denotiamo $u^{P+I} := u^P + u^I + 1 = |v|$. Abbiamo i seguenti risultati, condizionatamente a $(U^P = u^P, U^I = u^I)$,*

(i) *il vettore*

$$\mathbf{D} = \left[\log P_1 \quad \log P_2 \quad \dots \quad \log P_{u^{P+I}} \quad \log I_{u^I} \quad \dots \quad \log I_1 \right]' = \mathbf{R}\Omega_v \quad (5.14)$$

con \mathbf{R} matrice $u^{P+I} \times u^{P+I}$

$$\mathbf{R} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ 1 & 1 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \dots & 1 \end{bmatrix}, \quad (5.15)$$

segue una distribuzione $MSN_{u^{P+I}}$ di parametri $\mathbf{R}\boldsymbol{\mu}_v$, $\mathbf{R}\boldsymbol{\Sigma}_v^{1/2}$ e $\boldsymbol{\Delta}_v$;

(ii) *il valore atteso di P_t (con $t = 1, \dots, u^P + 1$) è*

$$\begin{aligned} & E[P_t | U^P = u^P, U^I = u^I] \\ &= 2^{u^{P+I}} \cdot \exp\left(\mathbf{R}_{[t,]} \boldsymbol{\mu}_v + 0.5 \mathbf{R}_{[t,]} (\boldsymbol{\Sigma}_v^{1/2}) (\boldsymbol{\Sigma}_v^{1/2})' \mathbf{R}'_{[t,]}\right) \cdot \prod_{j=1}^{u^{P+I}} \Phi\left(\frac{\Delta_{v,j} \left((\boldsymbol{\Sigma}_v^{1/2})' \mathbf{R}'_{[t,]}\right)_j}{\sqrt{1 + \Delta_{v,j}^2}}\right), \end{aligned} \quad (5.16)$$

dove $\mathbf{R}_{[t,]}$ è la t -esima riga di \mathbf{R} ;

il valore atteso di I_t (con $t = 1, \dots, u^I$) è

$$\begin{aligned} & E[I_t | U^P = u^P, U^I = u^I] \\ &= 2^{u^{P+I}} \cdot \exp\left(\mathbf{R}_{[t^*, \cdot]} \boldsymbol{\mu}_v + 0.5 \mathbf{R}_{[t^*, \cdot]} (\boldsymbol{\Sigma}_v^{1/2}) (\boldsymbol{\Sigma}_v^{1/2})' \mathbf{R}'_{[t^*, \cdot]}\right) \cdot \prod_{j=1}^{u^{P+I}} \Phi\left(\frac{\Delta_{v,j} \left((\boldsymbol{\Sigma}_v^{1/2})' \mathbf{R}'_{[t^*, \cdot]} \right)_j}{\sqrt{1 + \Delta_{v,j}^2}}\right), \end{aligned} \quad (5.17)$$

con $t^* = u^{P+I} - t + 1$.

(iii) la covarianza tra e^{D_i} e e^{D_j} (con $i, j = 1, \dots, u^{P+I}$) è

$$\begin{aligned} & Cov[e^{D_i}, e^{D_j} | U^P = u^P, U^I = u^I] \\ &= 2^{u^{P+I}} \cdot \exp\left(\mathbf{t}'_{i+j} \mathbf{R} \boldsymbol{\mu}_v + 0.5 \mathbf{t}'_{i+j} \mathbf{R} (\boldsymbol{\Sigma}_v^{1/2}) (\boldsymbol{\Sigma}_v^{1/2})' \mathbf{R}' \mathbf{t}_{i+j}\right) \cdot \prod_{n=1}^{u^{P+I}} \Phi\left(\frac{\Delta_{v,n} \left((\boldsymbol{\Sigma}_v^{1/2})' \mathbf{R}' \mathbf{t}_{i+j} \right)_n}{\sqrt{1 + \Delta_{v,n}^2}}\right) \\ &\quad - 2^{2(u^{P+I})} \cdot \exp\left(\mathbf{R}_{[i, \cdot]} \boldsymbol{\mu}_v + \mathbf{R}_{[j, \cdot]} \boldsymbol{\mu}_v + 0.5 \left((\mathbf{R}_{[i, \cdot]} (\boldsymbol{\Sigma}_v^{1/2}) (\boldsymbol{\Sigma}_v^{1/2})' \mathbf{R}'_{[i, \cdot]} + \mathbf{R}_{[j, \cdot]} (\boldsymbol{\Sigma}_v^{1/2}) (\boldsymbol{\Sigma}_v^{1/2})' \mathbf{R}'_{[j, \cdot]} \right)\right) \\ &\quad \times \prod_{n=1}^{u^{P+I}} \Phi\left(\frac{\Delta_{v,n} \left((\boldsymbol{\Sigma}_v^{1/2})' \mathbf{R}'_{[i, \cdot]} \right)_n}{\sqrt{1 + \Delta_{v,n}^2}}\right) \cdot \prod_{n=1}^{u^{P+I}} \Phi\left(\frac{\Delta_{v,n} \left((\boldsymbol{\Sigma}_v^{1/2})' \mathbf{R}'_{[j, \cdot]} \right)_n}{\sqrt{1 + \Delta_{v,n}^2}}\right), \end{aligned} \quad (5.18)$$

dove \mathbf{t}_{i+j} è la somma di due vettori colonna di dimensione u^{P+I} con '1' rispettivamente al posto i -esimo e j -esimo, '0' altrove.

Dimostrazione. Per brevità, nelle verifiche che seguono omettiamo il condizionamento a $(U^P = u^P, U^I = u^I)$, ma tutte le valutazioni probabilistiche devono essere intese in senso condizionato a tale evento.

(i) Per costruzione $\boldsymbol{\Omega}_v \sim MSN_{u^{P+I}}(\boldsymbol{\mu}_v, \boldsymbol{\Sigma}_v^{1/2}, \boldsymbol{\Delta}_v)$. Essendo (v. (5.3))

$$P_t = P_1 \cdot \prod_{j=1}^{t-1} \lambda_j, \quad t = 2, \dots, u^P + 1,$$

passando ai logaritmi otteniamo che

$$\log P_t = \log P_1 + \sum_{j=1}^{t-1} \log \lambda_j, \quad t = 2, \dots, u^P + 1. \quad (5.19)$$

Analogamente (v. (5.6)) otteniamo che

$$\log I_t = \log P_1 + \sum_{j=1}^{u^P} \log \lambda_j + \sum_{j=t}^{u^I} \log \gamma_j, \quad t = 1, \dots, u^I. \quad (5.20)$$

La funzione generatrice dei momenti di $\mathbf{D} = \mathbf{R}\boldsymbol{\Omega}_v$ è

$$\begin{aligned} m_{\mathbf{D}}(\mathbf{t}) &= m_{\mathbf{R}\boldsymbol{\Omega}_v}(\mathbf{t}) = E(e^{\mathbf{t}'\mathbf{R}\boldsymbol{\Omega}_v}) = m_{\boldsymbol{\Omega}_v}(\mathbf{R}'\mathbf{t}) \\ &= 2^{u^{P+I}} \cdot \exp\left(\mathbf{t}'\mathbf{R}\boldsymbol{\mu}_v + 0.5 \cdot \mathbf{t}'\mathbf{R}(\boldsymbol{\Sigma}_v^{1/2})(\boldsymbol{\Sigma}_v^{1/2})'\mathbf{R}'\mathbf{t}\right) \cdot \prod_{j=1}^{u^{P+I}} \Phi\left(\frac{\Delta_{v,j} \left((\boldsymbol{\Sigma}_v^{1/2})'\mathbf{R}'\mathbf{t}\right)_j}{\sqrt{1 + \Delta_{v,j}^2}}\right). \end{aligned}$$

Abbiamo trovato la funzione generatrice dei momenti di una distribuzione MSN con parametri aggiornati; poiché la funzione generatrice dei momenti caratterizza la distribuzione concludiamo che

$$\mathbf{D} \sim MSN_{u^{P+I}}\left(\mathbf{R}\boldsymbol{\mu}_v, \mathbf{R}\boldsymbol{\Sigma}_v^{1/2}, \boldsymbol{\Delta}_v\right).$$

(ii) Osserviamo che la (5.14) segue dalla (5.19), (5.20) con i seguenti passaggi

$$\mathbf{R}\boldsymbol{\Omega}_v = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ 1 & 1 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} \log P_1 \\ \log \lambda_1 \\ \vdots \\ \log \lambda_{u^P} \\ \log \gamma_{u^I} \\ \vdots \\ \log \gamma_1 \end{bmatrix} = \begin{bmatrix} \log P_1 \\ \log P_1 + \log \lambda_1 \\ \vdots \\ \log P_1 + \sum_{i=1}^{u^P} \log \lambda_i \\ \log P_1 + \sum_{i=1}^{u^P} \log \lambda_i + \log \gamma_{u^I} \\ \vdots \\ \log P_1 + \sum_{i=1}^{u^P} \log \lambda_i + \sum_{i=1}^{u^I} \log \gamma_i \end{bmatrix} = \begin{bmatrix} \log P_1 \\ \log P_2 \\ \vdots \\ \log P_{u^P+1} \\ \log I_{u^I} \\ \vdots \\ \log I_1 \end{bmatrix}.$$

Pertanto

$$\log P_t = \mathbf{R}_{[t,]} \boldsymbol{\Omega}_v, \quad t = 1, \dots, u^P + 1.$$

Condizionatamente a $U^P = u^P$, $U^I = u^I$ si ha

$$E(P_t) = E(e^{\log P_t}) = E(e^{\mathbf{R}_{[t,]} \boldsymbol{\Omega}_v}) = m_{\boldsymbol{\Omega}_v}(\mathbf{R}'_{[t,]})$$

da cui la tesi, per la (5.10).

Analogamente, notando che $t^* = u^{P+I}, \dots, 1$ per $t = 1, \dots, u^I$, si ha che

$$\log I_t = \mathbf{R}_{[t^*,]} \boldsymbol{\Omega}_v, \quad t = 1, \dots, u^I.$$

Perciò

$$E(I_t) = E(e^{\log I_t}) = E(e^{\mathbf{R}_{[t^*,]} \boldsymbol{\Omega}_v}) = m_{\boldsymbol{\Omega}_v}(\mathbf{R}'_{[t^*,]}).$$

(iii) Osserviamo preliminarmente che

$$e^{D_i} = e^{\mathbf{R}_{[i,]} \boldsymbol{\Omega}_v} = \begin{cases} P_i & i \leq u^{P+I} \\ I_i & \text{altrimenti} \end{cases}.$$

Condizionatamente a $U^P = u^P$ e $U^I = u^I$, per $i, j = 1, \dots, u^{P+I}$ abbiamo

$$\begin{aligned} Cov[e^{D_i}, e^{D_j}] &= E[e^{D_i+D_j}] - E[e^{D_i}]E[e^{D_j}] \\ &= E[e^{\mathbf{t}'_{i+j}\mathbf{R}\boldsymbol{\Omega}_v}] - E[e^{\mathbf{R}_{[i,]} \boldsymbol{\Omega}_v}]E[e^{\mathbf{R}_{[j,]} \boldsymbol{\Omega}_v}] \\ &= m_{\mathbf{R}\boldsymbol{\Omega}_v}(\mathbf{t}'_{i+j}) - m_{\boldsymbol{\Omega}_v}(\mathbf{R}'_{[i,]})m_{\boldsymbol{\Omega}_v}(\mathbf{R}'_{[j,]}). \end{aligned}$$

Sfruttando i risultati precedenti in (i) e (ii) segue la tesi. \square

Osservazione. Con la Proposizione 5.1 abbiamo ottenuto, condizionatamente a $(U^P = u^P, U^I = u^I)$, i valori attesi di P_t ($t = 1, \dots, u^P + 1$) e I_t ($t = 1, \dots, u^I$) e le covarianze tra pagamenti cumulati e *incurred losses*.

Analizziamo ora il caso in cui dati $U^P = u^P$ e $U^I = u^I$ vi sia una parte osservata del vettore aleatorio $\mathbf{D} = \mathbf{R}\boldsymbol{\Omega}_v$ e quindi di $\boldsymbol{\Omega}_v$.

Proposizione 5.2. *Dati $U^P = u^P$ e $U^I = u^I$, sia vo l'insieme degli indici corrispondenti alla parte osservata del vettore aleatorio $\boldsymbol{\Omega}_v$, vf l'insieme complementare di vo e $v' := \{vo, vf\}$. Sia t^P il numero di periodi con pagamenti parziali osservati e t^I il numero di periodi con modifiche osservate negli *incurred losses*, quindi $|v'| = t^P + t^I + 1$. Condizionatamente a $(U^P = u^P, U^I = u^I)$, il vettore aleatorio $\boldsymbol{\Omega}_{v'} = [\boldsymbol{\Omega}'_{vo} \quad \boldsymbol{\Omega}'_{vf}]'$ segue una distribuzione MSN di parametri $\boldsymbol{\mu}_{v'}$, $\boldsymbol{\Sigma}_{v'}^{1/2}$ e $\boldsymbol{\Delta}_{v'}$. Poniamo*

$$\boldsymbol{\mu}_{v'} := \begin{bmatrix} \tilde{\boldsymbol{\mu}}_o \\ \tilde{\boldsymbol{\mu}}_f \end{bmatrix}, \quad \boldsymbol{\Sigma}_{v'}^{1/2} := \begin{bmatrix} \tilde{\boldsymbol{\Sigma}}_{oo} & \mathbf{0} \\ \tilde{\boldsymbol{\Sigma}}_{fo} & \tilde{\boldsymbol{\Sigma}}_{ff} \end{bmatrix}, \quad \boldsymbol{\Delta}_{v'} := \begin{bmatrix} \tilde{\boldsymbol{\Delta}}_o \\ \tilde{\boldsymbol{\Delta}}_f \end{bmatrix},$$

dove: $\tilde{\boldsymbol{\mu}}_o$ e $\tilde{\boldsymbol{\Delta}}_o$ sono vettori colonna di dimensione $|vo|$; $\tilde{\boldsymbol{\mu}}_f$ e $\tilde{\boldsymbol{\Delta}}_f$ vettori colonna di dimensione $|vf|$; $\tilde{\boldsymbol{\Sigma}}_{oo}$ è una matrice triangolare inferiore di dimensione $|vo| \times |vo|$, con elementi diagonali positivi; $\tilde{\boldsymbol{\Sigma}}_{ff}$ è una matrice triangolare inferiore di dimensione $|vf| \times |vf|$, con elementi diagonali positivi. Siano $\boldsymbol{\omega}_{vo} = (\omega_1, \dots, \omega_{t^P}, \omega_{t^P+1}, \dots, \omega_{t^P+t^I+1})$ il valore osservato di $\boldsymbol{\Omega}_{vo}$, $\boldsymbol{\mu}^* := \tilde{\boldsymbol{\mu}}_f + \tilde{\boldsymbol{\Sigma}}_{fo}\tilde{\boldsymbol{\Sigma}}_{oo}^{-1}(\boldsymbol{\omega}_{vo} - \tilde{\boldsymbol{\mu}}_o)$, $\boldsymbol{\Sigma}^* := \tilde{\boldsymbol{\Sigma}}_{ff}$ e $\boldsymbol{\Delta}^* := \tilde{\boldsymbol{\Delta}}_f$.

Il valore atteso di P_{t^P+s} con $s = 1, \dots, u^P + 1 - t^P$, condizionatamente all'informazione osservata $\boldsymbol{\Omega}_{vo} = \boldsymbol{\omega}_{vo}$ e a $(U^P = u^P, U^I = u^I)$, è

$$\begin{aligned} & E[P_{t^P+s} | \boldsymbol{\Omega}_{vo} = \boldsymbol{\omega}_{vo}, U^P = u^P, U^I = u^I] \\ &= \exp\left(\sum_{j=1}^{t^P} \omega_j\right) \cdot 2^{|vf|} \exp\left(\mathbf{t}_s^* \boldsymbol{\mu}^* + 0.5 \mathbf{t}_s^* ((\boldsymbol{\Sigma}^*)^{1/2}) ((\boldsymbol{\Sigma}^*)^{1/2})' (\mathbf{t}_s^*)'\right) \\ & \cdot \prod_{j=1}^{|vf|} \Phi\left(\frac{\Delta_j^* (((\boldsymbol{\Sigma}^*)' (\mathbf{t}_s^*)')_j)}{\sqrt{1 + (\Delta_j^*)^2}}\right), \end{aligned} \quad (5.21)$$

con $\mathbf{t}_s^* := [1 \ \dots \ 1 \ 1 \ 0 \ \dots \ 0]$ vettore riga di dimensione $|vf|$ con 1 in posizione 1, ..., s e 0 altrove.

Il valore atteso di I_{t^I+s} con $s = 1, \dots, u^I - t^I$, condizionatamente all'informazione osservata $\boldsymbol{\Omega}_{vo} = \boldsymbol{\omega}_{vo}$ e a $(U^P = u^P, U^I = u^I)$, è

$$\begin{aligned} & E[I_{t^I+s} | \boldsymbol{\Omega}_{vo} = \boldsymbol{\omega}_{vo}, U^P = u^P, U^I = u^I] \\ &= \exp\left(\sum_{j=1}^{t^I} \omega_{|v'|-j}\right) \cdot 2^{|vf|} \exp\left(\mathbf{t}_s^{**} \boldsymbol{\mu}^* + 0.5 \mathbf{t}_s^{**} ((\boldsymbol{\Sigma}^*)^{1/2}) ((\boldsymbol{\Sigma}^*)^{1/2})' (\mathbf{t}_s^{**})'\right) \\ & \cdot \prod_{j=1}^{|vf|} \Phi\left(\frac{\Delta_j^* (((\boldsymbol{\Sigma}^*)^{1/2})' (\mathbf{t}_s^{**})')_j)}{\sqrt{1 + (\Delta_j^*)^2}}\right), \end{aligned} \quad (5.22)$$

con $\mathbf{t}_s^{**} := [0 \ \dots \ 0 \ 0 \ 1 \ \dots \ 1]$ vettore riga di dimensione $|vf|$ con 1 in posizione $|vf| - s + 1, \dots, |vf|$ e 0 altrove.

Dimostrazione. Poniamo

$$\Psi := \begin{bmatrix} \mathbf{1}_{|vo| \times |vo|} & \mathbf{0}_{|vo| \times |vf|} \\ -\tilde{\Sigma}_{fo} & \tilde{\Sigma}_{oo}^{-1} \\ \mathbf{1}_{|vf| \times |vf|} & \end{bmatrix}.$$

Abbiamo

$$\Psi \Omega_{v'} = \begin{bmatrix} \Omega_{vo} \\ -\tilde{\Sigma}_{fo} \tilde{\Sigma}_{oo}^{-1} \Omega_{vo} + \Omega_{vf} \end{bmatrix} = \begin{bmatrix} \Omega_{vo} \\ \Omega_{vf}^* \end{bmatrix}$$

e, condizionatamente a $(U^P = u^P, U^I = u^I)$,

$$m_{\Psi \Omega_{v'}}(\mathbf{t}) = m_{\Omega_{v'}}(\Psi' \mathbf{t}).$$

Poiché, per la Proposizione 5.1 (i), $\Omega_{v'}$ segue una distribuzione $MSN(\boldsymbol{\mu}_{v'}, \Sigma_{v'}^{1/2}, \Delta_{v'})$, dalla corrispondenza tra distribuzione di probabilità e funzione generatrice dei momenti abbiamo che $\Psi \Omega_{v'} \sim MSN(\Psi \boldsymbol{\mu}_{v'}, \Psi \Sigma_{v'}^{1/2}, \Delta_{v'})$, con

$$\Psi \boldsymbol{\mu}_{v'} = \begin{bmatrix} \tilde{\boldsymbol{\mu}}_o \\ \tilde{\boldsymbol{\mu}}_f - \tilde{\Sigma}_{fo} \tilde{\Sigma}_{oo}^{-1} \tilde{\boldsymbol{\mu}}_o \end{bmatrix}, \quad \Psi \Sigma_{v'}^{1/2} = \begin{bmatrix} \tilde{\Sigma}_{oo} & \mathbf{0} \\ \mathbf{0} & \tilde{\Sigma}_{ff} \end{bmatrix}, \quad \Delta_{v'} = \begin{bmatrix} \tilde{\Delta}_o \\ \tilde{\Delta}_f \end{bmatrix}.$$

La densità di $\Psi \Omega_{v'}$ condizionata a $(U^P = u^P, U^I = u^I)$ è

$$\begin{aligned} & \frac{2^{|vo|}}{\det(\tilde{\Sigma}_{oo})^{1/2}} \cdot \phi_{|vo|} \left(\tilde{\Sigma}_{oo}^{-1/2} (\boldsymbol{\omega}_{vo} - \tilde{\boldsymbol{\mu}}_o) \right) \prod_{j=1}^{|vo|} \Phi \left(\Delta_{v',j} e'_j \tilde{\Sigma}_{oo}^{-1/2} (\boldsymbol{\omega}_{vo} - \tilde{\boldsymbol{\mu}}_o) \right) \\ & \cdot \frac{2^{|vf|}}{\det(\tilde{\Sigma}_{ff})^{1/2}} \cdot \phi_{|vf|} \left(\tilde{\Sigma}_{ff}^{-1/2} \left(\boldsymbol{\omega}_{vf}^* - \left(\tilde{\boldsymbol{\mu}}_f - \tilde{\Sigma}_{fo} \tilde{\Sigma}_{oo}^{-1} \tilde{\boldsymbol{\mu}}_o \right) \right) \right) \\ & \cdot \prod_{j=|vo|+1}^{|v|} \Phi \left(\Delta_{v',j} e'_j \tilde{\Sigma}_{ff}^{-1/2} \left(\boldsymbol{\omega}_{vf}^* - \left(\tilde{\boldsymbol{\mu}}_f - \tilde{\Sigma}_{fo} \tilde{\Sigma}_{oo}^{-1} \tilde{\boldsymbol{\mu}}_o \right) \right) \right). \end{aligned} \quad (5.23)$$

Riconosciamo che la precedente espressione è il prodotto di due densità

$$MSN_{|vo|} \left(\tilde{\boldsymbol{\mu}}_o, \tilde{\Sigma}_{oo}^{1/2}, \tilde{\Delta}_o \right) \times MSN_{|vf|} \left(\tilde{\boldsymbol{\mu}}_f - \tilde{\Sigma}_{fo} \tilde{\Sigma}_{oo}^{-1} \tilde{\boldsymbol{\mu}}_o, \tilde{\Sigma}_{ff}^{1/2}, \tilde{\Delta}_f \right).$$

Quindi condizionatamente a $(U^P = u^P, U^I = u^I)$, le componenti Ω_{vo} e Ω_{vf}^* sono stocasticamente indipendenti. Se consideriamo la densità di Ω_{vf}^* condizionata anche a $\Omega_{vo} = \boldsymbol{\omega}_{vo}$

e sostituiamo l'espressione $\boldsymbol{\omega}_{vf}^* = \boldsymbol{\omega}_{vf} - \tilde{\boldsymbol{\Sigma}}_{fo}\tilde{\boldsymbol{\Sigma}}_{oo}^{-1}\boldsymbol{\omega}_{vo}$, otteniamo

$$\begin{aligned} & \frac{2^{|vf|}}{\det(\tilde{\boldsymbol{\Sigma}}_{ff})^{1/2}} \cdot \phi_{|vf|} \left(\tilde{\boldsymbol{\Sigma}}_{ff}^{-1/2} (\boldsymbol{\omega}_{vf} - \tilde{\boldsymbol{\Sigma}}_{fo}\tilde{\boldsymbol{\Sigma}}_{oo}^{-1}\boldsymbol{\omega}_{vo} - \tilde{\boldsymbol{\mu}}_f + \tilde{\boldsymbol{\Sigma}}_{fo}\tilde{\boldsymbol{\Sigma}}_{oo}^{-1}\tilde{\boldsymbol{\mu}}_o) \right) \\ & \cdot \prod_{j=|vo|+1}^{|v|} \Phi \left(\Delta_{v',j} e'_j \tilde{\boldsymbol{\Sigma}}_{ff}^{-1/2} (\boldsymbol{\omega}_{vf} \tilde{\boldsymbol{\Sigma}}_{fo}\tilde{\boldsymbol{\Sigma}}_{oo}^{-1} - \boldsymbol{\omega}_{vo} - \tilde{\boldsymbol{\mu}}_f + \tilde{\boldsymbol{\Sigma}}_{fo}\tilde{\boldsymbol{\Sigma}}_{oo}^{-1}\tilde{\boldsymbol{\mu}}_o) \right) \\ & = \frac{2^{|vf|}}{\det(\tilde{\boldsymbol{\Sigma}}_{ff})^{1/2}} \cdot \phi_{|vf|} \left(\tilde{\boldsymbol{\Sigma}}_{ff}^{-1/2} (\boldsymbol{\omega}_{vf} - \tilde{\boldsymbol{\mu}}_f + \tilde{\boldsymbol{\Sigma}}_{fo}\tilde{\boldsymbol{\Sigma}}_{oo}^{-1}(\tilde{\boldsymbol{\mu}}_o - \boldsymbol{\omega}_{vo})) \right) \\ & \cdot \prod_{j=|vo|+1}^{|v|} \Phi \left(\Delta_{v',j} e'_j \tilde{\boldsymbol{\Sigma}}_{ff}^{-1/2} (\boldsymbol{\omega}_{vf} - \tilde{\boldsymbol{\mu}}_f + \tilde{\boldsymbol{\Sigma}}_{fo}\tilde{\boldsymbol{\Sigma}}_{oo}^{-1}(\tilde{\boldsymbol{\mu}}_o - \boldsymbol{\omega}_{vo})) \right). \end{aligned}$$

Pertanto, ricaviamo che condizionatamente a $(\boldsymbol{\Omega}_{vo} = \boldsymbol{\omega}_{vo}, U^P = u^P, U^I = u^I)$,

$$\boldsymbol{\Omega}_{vf} \sim MSN(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^{*1/2}, \boldsymbol{\Delta}_f^*),$$

con $\boldsymbol{\mu}^* = \tilde{\boldsymbol{\mu}}_f + \tilde{\boldsymbol{\Sigma}}_{fo}\tilde{\boldsymbol{\Sigma}}_{oo}^{-1}(\boldsymbol{\omega}_{vo} - \tilde{\boldsymbol{\mu}}_o)$, $\boldsymbol{\Sigma}^{*1/2} = \tilde{\boldsymbol{\Sigma}}_{ff}^{1/2}$ e $\boldsymbol{\Delta}_f^* = \tilde{\boldsymbol{\Delta}}_f$. Condizionatamente a $(\boldsymbol{\Omega}_{vo} = \boldsymbol{\omega}_{vo}, U^P = u^P, U^I = u^I)$ abbiamo allora che

$$\begin{aligned} E(P_{t^P+s}) & = E(e^{\log P_{t^P+s}}) = E(e^{\log P_{t^P} + t_s^* \boldsymbol{\Omega}_{vf}}) = e^{\log P_{t^P}} \cdot E(e^{t_s^* \boldsymbol{\Omega}_{vf}}) \\ & = \exp \left(\sum_{j=1}^{t^P} \omega_j \right) \cdot m_{\boldsymbol{\Omega}_{vf}} \left((t_s^*)' \right) = \\ & = \exp \left(\sum_{j=1}^{t^P} \omega_j \right) \cdot 2^{|vf|} \cdot \exp \left(t_s^* \boldsymbol{\mu}^* + 0.5 t_s^* (\boldsymbol{\Sigma}^{*1/2}) (\boldsymbol{\Sigma}^{*1/2})' (t_s^*)' \right) \\ & \cdot \prod_{j=1}^{|vf|} \Phi \left(\frac{\Delta_{f,j}^* \left((\boldsymbol{\Sigma}^{*1/2})' (t_s^*)' \right)_j}{\sqrt{1 + \Delta_j^{*2}}} \right), \end{aligned}$$

cioè la tesi. Analogamente si ha

$$\begin{aligned} E(I_{t^I+s}) & = E(e^{\log I_{t^I+s}}) = E(e^{\log I_{t^I} + t_s^{**} \boldsymbol{\Omega}_{vf}}) = e^{\log I_{t^I}} \cdot E(e^{t_s^{**} \boldsymbol{\Omega}_{vf}}) \\ & = \exp \left(\sum_{j=1}^{t^I} \omega_{|v'|-j} \right) \cdot m_{\boldsymbol{\Omega}_{vf}} \left((t_s^{**})' \right) = \\ & = \exp \left(\sum_{j=1}^{t^I} \omega_{|v'|-j} \right) \cdot 2^{|vf|} \cdot \exp \left(t_s^{**} \boldsymbol{\mu}^* + 0.5 t_s^{**} (\boldsymbol{\Sigma}^{*1/2}) (\boldsymbol{\Sigma}^{*1/2})' (t_s^{**})' \right) \\ & \cdot \prod_{j=1}^{|vf|} \Phi \left(\frac{\Delta_{f,j}^* \left((\boldsymbol{\Sigma}^{*1/2})' (t_s^{**})' \right)_j}{\sqrt{1 + \Delta_j^{*2}}} \right). \end{aligned}$$

□

Osservazione. Occorre scegliere opportuni sottoinsiemi vo e vf di v . Ad esempio, se si dispone delle osservazioni su P_1, P_2, P_3 e I_1, I_2 , si ha

$$vo = \{1, 2, 3, u^{P+I-1}, u^{P+I}\}, \quad vf = \{4, 5, \dots, u^{P+I-3}, u^{P+I-2}\}.$$

Nella seguente proposizione otteniamo infine un'espressione per le riserve IBNR, RBNP e RBNS:

Proposizione 5.3 (Best estimates per le riserve IBNR, RBNP e RBNS). *Sia \mathcal{I}_τ l'informazione congiunta disponibile per tutti i sinistri del dataset al tempo τ e $\mathcal{I}_{k,\tau}$ l'informazione disponibile per il sinistro k . Quest'ultima è la parte osservata dello sviluppo del sinistro k , ovvero Ω_{vo_k} .*

(i) *I valori attesi delle riserve IBNR e RBNP sono, rispettivamente,*

$$E[IBNR|\mathcal{I}_\tau] = E[K_{IBNR}] \cdot E_{U^P, U^I}[E[P_{U^P+1}|U^P, U^I]]$$

e

$$E[RBNP|\mathcal{I}_\tau] = k_{RBNP} \cdot E_{U^P, U^I}[E[P_{U^P+1}|U^P, U^I]],$$

dove K_{IBNR} è il numero aleatorio di sinistri IBNR⁴, $\mathbf{E}(P_{U^P+1}|U^P, U^I)$ è il valore atteso condizionato a (U^P, U^I) del costo ultimo per un sinistro, ottenibile dalla (5.16), k_{RBNP} è il numero osservato di sinistri aperti senza pagamenti e senza una stima dell'inventario iniziale.

(ii) *Il valore atteso del costo ultimo RBNS è*

$$E[RBNS|\mathcal{I}_\tau] = \sum_{k=1}^{\mathcal{K}_{RBNS}} E_{U^P, U^I} \left[E[P_{k, U_k^P+1} | \mathcal{I}_{k,\tau} = \omega_{vo_k}, U_k^P, U_k^I] \right],$$

dove la sommatoria è relativa a tutti i sinistri RBNS del dataset. \mathcal{K}_{RBNS} è il numero osservato di sinistri RBNS e $\mathbf{E}(P_{k, U_k^P+1} | \mathcal{I}_{k,\tau} = \omega_{vo_k}, U_k^P, U_k^I)$ è il valore del costo ultimo atteso condizionato per il sinistro k -esimo, ottenibile dalla (5.21). La stima della riserva RBNS si ottiene sottraendo il pagamento cumulato osservato più recente.

Concludiamo il capitolo osservando che per applicare il modello occorre disporre di dati dettagliati, come si è detto nel Paragrafo 5.1. Dai dati si stimano anche i parametri del modello. L'approccio di stima dei parametri proposto dagli Autori prevede di utilizzare il metodo della massima verosimiglianza per la stima delle distribuzioni dei numeri aleatori $U_k^P, U_k^I, Q_k^P, Q_k^I, N_{k,j}^P, N_{k,j}^I$. La stima degli altri parametri è ottenuta con un procedimento iterativo.

⁴ Per valutare $\mathbf{E}(k_{IBNR})$ occorre introdurre una ipotesi probabilistica. Gli Autori propongono di assegnare la valutazione con un approccio analogo a quello seguito per assegnare la distribuzione del numero di sinistri del periodo di origine i , K_i (v. Paragrafo 5.2.2). Assegnano dunque a $K_{IBNR,i}$, numero di sinistri IBNR del periodo di origine i , una distribuzione di Poisson con parametro $\theta\omega(i)(1 - F_1(r_i^* - 1; \nu))$.

Capitolo 6

Valutazioni numeriche

In questo capitolo è presentata un'applicazione numerica che ha l'obiettivo di confrontare le valutazioni ottenute con i modelli Chain Ladder, Munich Chain Ladder e Paid-incurred Chain su un medesimo insieme di dati. Come nelle esemplificazioni numeriche dei capitoli precedenti, è stato utilizzato il linguaggio R e il pacchetto `ChainLadder` per elaborare il confronto. Nell'Appendice A è riportato il codice implementato a tale scopo.

6.1 R e il pacchetto `ChainLadder`

R è un linguaggio di programmazione open source e un ambiente di sviluppo ampiamente utilizzato per lo sviluppo di software statistico. E' disponibile liberamente sotto la GNU General Public License¹. La potenzialità di R si estende ulteriormente attraverso l'uso di pacchetti creati dagli utenti e disponibili con licenza GPL sulla rete di archivi CRAN (Comprehensive R Archive Network).

Il pacchetto che è stato maggiormente utilizzato in questo lavoro di tesi è stato il pacchetto `ChainLadder`. La prima versione è stata creata da Markus Gesmann nel 2008; gli Autori della versione attuale (0.2.0) sono Alessandro Carrato, Markus Gesmann, Dan Murphy, Mario Wüthrich e Wayne Zhang. Tale pacchetto fornisce vari metodi statistici che sono comunemente usati per la stima delle riserve sinistri nei rami danni. Sono ad esempio implementati il modello Chain Ladder di Mack, il modello Munich Chain Ladder, modelli che fanno uso di GLM e della simulazione bootstrap.

Le funzioni che sono state principalmente utilizzate per le valutazioni numeriche presenti nella tesi sono la funzione `MackChainLadder`, la quale implementa il modello di Mack, e la funzione `MunichChainLadder`, che implementa il modello Munich Chain Ladder. La funzione `MackChainLadder` è stata utilizzata con la seguente chiamata

```
MackChainLadder(Triangle, est.sigma="log-linear"),
```

con i seguenti input:

¹La GNU General Public License è una licenza per il software libero, che garantisce agli utenti (individui, organizzazioni, compagnie) la libertà di usare, studiare, condividere e modificare il software.

- `Triangle` il triangolo di run-off dei pagamenti cumulati o degli *incurred losses*;
- `est.sigma` il metodo di stima per il parametro σ_I ; "log-linear" per utilizzare una regressione log-lineare.

Tra gli oggetti forniti in output dalla funzione sono stati utilizzati i seguenti:

- `FullTriangle` il triangolo completo con le previsioni date dal metodo CL;
- `f` la stima dei fattori di sviluppo dei pagamenti o degli *incurred losses*;
- `Total.Mack.S.E` la radice quadrata della stima del msep della riserva di portafoglio.

La funzione `MunichChainLadder` è stata utilizzata con la seguente chiamata

```
MunichChainLadder(Paid, Incurred, est.sigmaP = "log-linear",  
                  est.sigmaI = "log-linear"),
```

con i seguenti input:

- `Paid` il triangolo di run-off dei pagamenti cumulati;
- `Incurred` il triangolo di run-off degli *incurred losses*;
- `est.sigmaP` il metodo di stima per il parametro σ_I^P ;
- `est.sigmaI` il metodo di stima per il parametro σ_I^I .

Tra gli oggetti forniti in output dalla funzione sono stati utilizzati i seguenti:

- `MCLPaid` il triangolo dei pagamenti cumulati completo con le previsioni date dal metodo MCL;
- `MCLIncurred` il triangolo degli *incurred losses* completo con le previsioni date dal metodo MCL;
- `lambdaP` la stima del coefficiente λ^P del metodo MCL;
- `lambdaI` la stima del coefficiente λ^I del metodo MCL.

Per una descrizione completa della chiamata delle due funzioni e dei relativi input e output si veda la documentazione del pacchetto².

Non è invece presente nel pacchetto una funzione relativa al modello Paid-incurred Chain. Per tale motivo ho implementato una funzione per tale modello nel caso particolare di distribuzioni a priori non informative; essa è stata accettata dagli Autori del pacchetto e sarà inclusa in esso a partire dalla versione 0.2.1.

²<http://cran.r-project.org/web/packages/ChainLadder/ChainLadder.pdf>.

6.2 Confronto numerico tra i modelli

Per il confronto numerico consideriamo i dati di un portafoglio americano di assicurazioni di responsabilità civile auto, per autoveicoli adibiti ad uso privato, della compagnia Celina Insurance Group³. Abbiamo a disposizione i triangoli di run-off dei pagamenti e degli *incurred losses* con $I = 9$. Gli anni di origine vanno dal 1988 al 1997. Nella Tabella 6.1 è riportato il triangolo dei pagamenti cumulati, mentre nella Tabella 6.2 è riportato il triangolo degli *incurred losses*.

Tabella 6.1: triangolo dei pagamenti cumulati.

	$P_{i,j}$									
	0	1	2	3	4	5	6	7	8	9
0	4,339	9,617	11,584	12,001	12,640	12,966	13,058	13,080	13,182	13,183
1	4,445	9,090	11,125	11,815	12,027	12,078	12,626	12,626	12,627	
2	6,123	10,564	12,290	13,579	14,063	14,412	14,488	14,497		
3	5,464	9,537	11,481	12,335	12,699	12,721	12,725			
4	5,113	8,527	10,344	11,381	11,742	12,036				
5	6,540	11,810	14,533	14,998	15,890					
6	5,781	9,519	10,929	11,356						
7	4,356	6,779	8,199							
8	5,250	8,967								
9	5,743									

Nella Tabella 6.3 sono riportati i fattori di sviluppo del metodo della catena, mentre nella Tabella 6.4 sono riportati i fattori di sviluppo aggiornati dei pagamenti del modello MCL. Non sono stati riportati i fattori di sviluppo aggiornati degli *incurred losses* del modello MCL poiché, se si riportano i valori con 4 cifre decimali, non vi sono modifiche rispetto ai fattori di sviluppo del modello CL. Ciò è dovuto a un coefficiente λ^I quasi nullo e pari a -0.0665.

Nella Tabella 6.5 sono riportate le stime delle riserve per anno di origine e delle riserve di portafoglio per ciascun metodo. Si notano i seguenti asserti.

- Applicando il metodo Chain Ladder, le stime della riserva di portafoglio usando i pagamenti e usando gli *incurred losses* differiscono sensibilmente, come è lecito aspettarsi sulla base delle osservazioni fatte nel Capitolo 2.
- Applicando il metodo Munich Chain Ladder tale differenza si riduce, ma il metodo corregge verso il basso entrambe le stime.
- La stima di portafoglio ottenuta con il metodo Paid-incurred Chain è in questo caso compresa tra la stima *paid* e la stima *incurred* del metodo Munich Chain Ladder.

³Fonte http://www.casact.org/research/index.cfm?fa=loss_reserves_data.

Tabella 6.2: triangolo degli *incurred losses*.

	$I_{i,j}$									
	0	1	2	3	4	5	6	7	8	9
0	12,705	12,990	12,728	12,730	13,020	13,077	13,098	13,181	13,178	13,178
1	11,245	11,379	12,109	12,220	12,344	12,541	12,624	12,637	12,635	
2	13,223	13,045	13,666	14,407	14,448	14,516	14,499	14,491		
3	12,254	12,001	12,345	12,892	12,890	12,752	12,717			
4	11,882	11,582	11,815	12,043	12,146	12,187				
5	14,906	15,174	15,981	16,062	16,172					
6	12,532	12,481	12,530	12,387						
7	10,639	9,879	9,947							
8	11,804	11,612								
9	10,455									

Tabella 6.3: fattori di sviluppo del metodo CL.

	j								
	0	1	2	3	4	5	6	7	8
f_j^P	1.7804	1.1994	1.0629	1.0388	1.0165	1.0138	1.0008	1.0040	1.0001
f_j^I	0.9906	1.0263	1.0172	1.0083	1.0035	1.0010	1.0022	0.9998	1.0000

Tabella 6.4: fattori di sviluppo aggiornati dei pagamenti del metodo MCL.

	j								
	0	1	2	3	4	5	6	7	8
$f_{0,j}^P$	1.7812	1.1994	1.0629	1.0388	1.0165	1.0138	1.0008	1.0040	1.0001
$f_{1,j}^P$	1.7807	1.1993	1.0629	1.0387	1.0165	1.0139	1.0008	1.0040	1.0001
$f_{2,j}^P$	1.7802	1.1993	1.0629	1.0388	1.0165	1.0138	1.0008	1.0040	1.0001
$f_{3,j}^P$	1.7803	1.1993	1.0629	1.0388	1.0164	1.0138	1.0008	1.0040	1.0001
$f_{4,j}^P$	1.7804	1.1994	1.0630	1.0388	1.0165	1.0138	1.0008	1.0040	1.0001
$f_{5,j}^P$	1.7803	1.1994	1.0629	1.0388	1.0165	1.0138	1.0008	1.0040	1.0001
$f_{6,j}^P$	1.7802	1.1994	1.0630	1.0388	1.0165	1.0138	1.0008	1.0040	1.0001
$f_{7,j}^P$	1.7806	1.1995	1.0631	1.0389	1.0166	1.0138	1.0008	1.0040	1.0001
$f_{8,j}^P$	1.7803	1.1994	1.0629	1.0388	1.0165	1.0138	1.0008	1.0040	1.0001
$f_{9,j}^P$	1.7796	1.1992	1.0628	1.0387	1.0164	1.0138	1.0008	1.0040	1.0001

Tabella 6.5: stime delle riserve.

i	CL paid	CL incurred	MCL paid	MCL incurred	PIC
1	0.96	8	10.40	8.14	7.86
2	59.19	-8.8	-29.79	-9.66	5.26
3	61.82	17.4	-1.32	14.05	31.64
4	225.36	187.3	172.75	186.06	198.97
5	564.54	386.5	356.62	374.54	431.39
6	859.56	1,214.3	1,228.28	1,241.99	1,052.31
7	1,175.70	2,068.7	2,132.59	2,143.78	1,641.53
8	3,330.06	3,334.5	3,321.17	3,334.63	3,347.77
9	8,278.93	5,228.5	5,026.18	5,038.12	5,533.97
Totale	14,556.11	12,436.32	12,216.88	12,331.65	12,250.70

Dai risultati ottenuti è possibile ritenere che, per questi dati, il metodo CL tradizionale, applicato ai *paid*, fornisce una sovrastima della riserva sinistri.

Nella Tabella 6.6 sono infine riportate le stime del msep. Segnaliamo innanzitutto che per il metodo Munich Chain Ladder non sono disponibili espressioni in forma chiusa per il calcolo del msep; questo rimane ad oggi un problema aperto. L'errore di previsione relativo, ottenuto come rapporto tra $\text{msep}^{1/2}$ e la stima della riserva, è pari a 0.076 nel modello PIC. Un valore piuttosto vicino al valore del metodo CL applicato agli *incurred losses* che è 0.082, mentre è decisamente più basso rispetto al valore del CL tradizionale che è 0.152. Notiamo che l'incertezza legata alla stima delle riserve di portafoglio si riduce utilizzando il metodo PIC rispetto al metodo CL.

Tabella 6.6: stime del msep.

	\hat{R}	$\text{msep}^{1/2}$	$\text{msep}^{1/2}/\hat{R}$
CL paid	14,556.11	2,216.50	0.152
CL incurred	12,436.32	1,020.14	0.082
MCL paid	12,216.88	Non disponibile	-
MCL incurred	12,331.65	Non disponibile	-
PIC	12,250.70	929.04	0.076

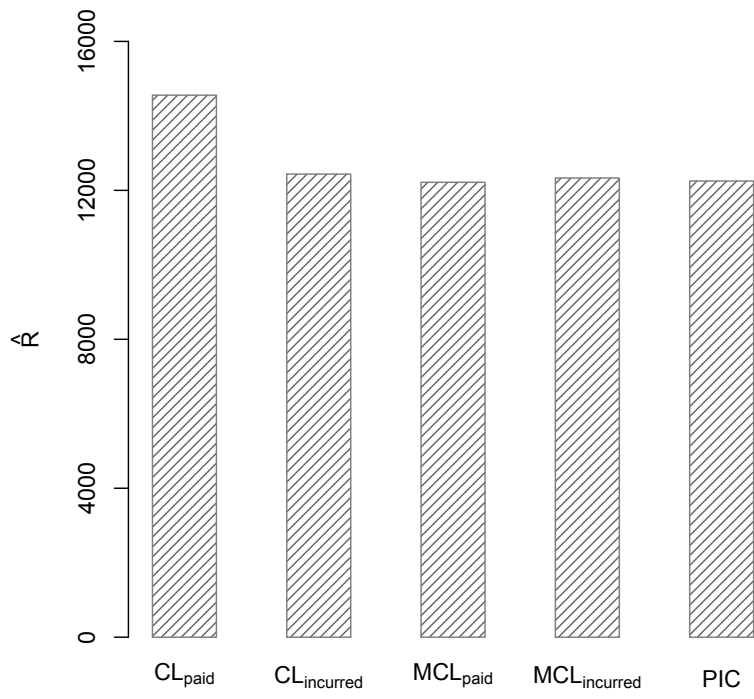


Figura 6.1: Riserve di portafoglio.

Appendice A

Codice R

In questa appendice viene riportato il codice R utilizzato per le valutazioni numeriche del Capitolo 6.

```
1 # Confronto tra i metodi Chain Ladder, Munich Chain Ladder
2 # e Paid Incurred Chain
3
4 library(ChainLadder)
5
6 # La seguente parte di codice per estrarre i dati dal file .csv
7 # e' stata scaricata da
8 # http://www.casact.org/research/index.cfm?fa=loss\_reserves\_data
9
10 # -----
11 a=read.csv("ppauto_pos.csv",header=TRUE)
12
13 grp.code=unique(a$GRCODE)
14 #
15 # function to get Schedule P triangle data given ins group
16 ins.line.data=function(g.code){
17   b=subset(a,a$GRCODE==g.code)
18   name=b$GRNAME
19   grpcode=b$GRCODE
20   ay=b$AccidentYear
21   dev=b$DevelopmentLag
22
23   cum_incloss=b[,6]
24   cum_pdloss=b[,7]
25   bulk_loss=b[,8]
26   dir_premium=b[,9]
27   ced_premium=b[,10]
28   net_premium=b[,11]
29   single=b[,12]
30   posted_reserve97=b[,13]
31
```

```
32 # get incremental paid losses - assume data is sorted by ay and
    dev
33 inc_pdloss=numeric(0)
34 for (i in unique(ay)){
35   s=(ay==i)
36   pl=c(0,cum_pdloss[s])
37   ndev=length(pl)-1
38   il=rep(0,ndev)
39   for (j in 1:ndev){
40     il[j]=pl[j+1]-pl[j]
41   }
42   inc_pdloss=c(inc_pdloss,il)
43 }
44 data.out=data.frame(name,grpcode,ay,dev,net_premium,dir_premium,
    ced_premium,
45   cum_pdloss,cum_incloss,bulk_loss,inc_pdloss,single,posted_
    reserve97)
46 return(data.out)
47 }
48 # -----
49
50 # seleziono la compagnia
51 data<-ins.line.data(grp.code[3]) # 3 = Celina Mut Grp
52
53 # triangolo superiore
54 data.up=subset(data,ay+dev<=1998)
55
56 tri <- as.triangle(data.up,origin="ay",dev="dev","cum_pdloss")
57 tri2 <- as.triangle(data.up,origin="ay",dev="dev","cum_incloss")
58
59 tri <- as.matrix(tri)
60 dimnames(tri) <- list(origin=0:(ncol(tri)-1), dev=1:ncol(tri))
61 tri <- as.triangle(tri)
62
63 tri2 <- as.matrix(tri2)
64 dimnames(tri2) <- list(origin=0:(ncol(tri2)-1), dev=1:ncol(tri2))
65 tri2 <- as.triangle(tri2)
66
67 J <- ncol(tri) # numero di colonne, comune a entrambi i triangoli
68 diagP <- diag(tri[J:1, 1:J])[(J-1):1]
69
70
71
72 # ----- Chain Ladder (modello di Mack) -----
73
74 # applico il metodo al triangolo dei pagamenti cumulati
75 CL.P <- MackChainLadder(tri, est.sigma = "log-linear")
76
77 # estraggo vettore costi ultimi paid
```

```

78 CL.Pult <- CL.P$FullTriangle[, J]
79 CL.Pult <- CL.Pult[-1]
80 # calcolo riserve per anno di origine paid
81 CL.PRis <- CL.Pult - diagP
82 # calcolo riserva di portafoglio paid
83 CL.PRisTot <- sum(CL.PRis)
84
85
86 # applico il metodo al triangolo degli incurred losses
87 CL.I <- MackChainLadder(tri2, est.sigma="log-linear")
88
89 # estraggo vettore costi ultimi incurred
90 CL.Iult <- CL.I$FullTriangle[, J]
91 CL.Iult <- CL.Iult[-1]
92 # calcolo riserve per anno di origine incurred
93 CL.IRis <- CL.Iult - diagP
94 # calcolo riserva di portafoglio incurred
95 CL.IRisTot <- sum(CL.IRis)
96
97 # calcolo MSEP
98 CL.P.se <- CL.P$Total.Mack.S.E
99 CL.I.se <- CL.I$Total.Mack.S.E
100
101
102 # ----- Munich Chain Ladder -----
103
104 MCL <- MunichChainLadder(tri, tri2, est.sigmaP="log-linear", est.
      sigmaI="log-linear")
105
106 # estraggo vettore costi ultimi paid
107 MCL.Pult <- MCL$MCLPaid[, J]
108 MCL.Pult <- MCL.Pult[-1]
109 # calcolo riserve per anno di origine paid
110 MCL.PRis <- MCL.Pult - diagP
111 # calcolo riserva di portafoglio paid
112 MCL.PRisTot <- sum(MCL.PRis)
113
114 # estraggo vettore costi ultimi incurred
115 MCL.Iult <- MCL$MCLIncurred[, J]
116 MCL.Iult <- MCL.Iult[-1]
117 # calcolo riserve per anno di origine paid
118 MCL.IRis <- MCL.Iult - diagP
119 # calcolo riserva di portafoglio paid
120 MCL.IRisTot <- sum(MCL.IRis)
121
122 # calcolo i residui dei rapporti Q
123 Q <- MCL$MCLPaid/MCL$MCLIncurred
124 EQ <- c()
125 VarQ <- matrix(data=NA, nrow=J, ncol=J)

```

```

126 ResQ <- matrix(data=NA, nrow=J, ncol=J)
127 for (i in 1:J) {
128   for (j in 1:J) {
129     EQ[j] <- mean(Q[,j])
130     somma <- 0
131     for (h in 1:(J-j+1)) {
132       somma <- somma + MCL$MCLIncurred[h,j] * (Q[h,j] - EQ[j])^2
133     }
134     VarQ[i,j] <- somma/((J-j+1) * MCL$MCLIncurred[i,j])
135     ResQ[i,j] <- (Q[i,j] - EQ[j])/sqrt(VarQ[i,j])
136   }
137 }
138
139 # calcolo i residui dei rapporti Qiniv
140 Qiniv <- MCL$MCLIncurred/MCL$MCLPaid
141 EQiniv <- c()
142 VarQiniv <- matrix(data=NA, nrow=J, ncol=J)
143 ResQiniv <- matrix(data=NA, nrow=J, ncol=J)
144 for (i in 1:J) {
145   for (j in 1:J) {
146     EQiniv[j] <- mean(Qiniv[,j])
147     somma <- 0
148     for (h in 1:(J-j+1)) {
149       somma <- somma + MCL$MCLPaid[h,j] * (Qiniv[h,j] - EQiniv[j])^2
150     }
151     VarQiniv[i,j] <- somma/((J-j+1) * MCL$MCLPaid[i,j])
152     ResQiniv[i,j] <- (Qiniv[i,j] - EQiniv[j])/sqrt(VarQiniv[i,j])
153   }
154 }
155
156 # calcolo i link ratios modificati
157 MCL.fP <- matrix(data=NA, nrow=J, ncol=J)
158 for (i in 1:J) {
159   for (j in 1:J) {
160     MCL.fP[i,j] <- CL.P$f[j] + (MCL$lambdaP$coefficients * CL.P$f.
161       se[j] * ResQiniv[i,j])/sqrt(MCL$MCLIncurred[i,j])
162   }
163 }
164 MCL.fI <- matrix(data=NA, nrow=J, ncol=J)
165 for (i in 1:J) {
166   for (j in 1:J) {
167     MCL.fI[i,j] <- CL.I$f[j] + (MCL$lambdaI$coefficients * CL.I$f.
168       se[j] * ResQ[i,j])/sqrt(MCL$MCLPaid[i,j])
169   }
170 }
171 # ----- Paid Incurred Chain -----
172 # tale metodo non e' presente nel pacchetto ChainLadder 0.2.0

```

```

173 # e' stato percio' implementato nel caso particolare di
174 # distribuzioni a priori non informative
175
176 # calcolo triangolo dei log(P_{i,j} / P_{i,j-1})
177 fP <- matrix(data=NA, nrow=J, ncol=J)
178 for (i in 1:J) {
179   fP[i, 1] <- log(tri[i, 1])
180   if (i == J) {
181     break
182   }
183   for (j in 1:(J-i)) {
184     fP[i, j+1] <- log(tri[i, j+1] / tri[i, j])
185   }
186 }
187 fP <- as.triangle(fP)
188
189 # calcolo triangolo dei log(I_{i,j} / I_{i,j+1})
190 fI <- matrix(data=NA, nrow=J, ncol=J)
191 for (i in 1:(J-1)) {
192   for (j in 1:(J-i)) {
193     fI[i, j] <- log(tri2[i, j] / tri2[i, j+1])
194   }
195 }
196 fI <- as.triangle(fI)
197
198 # stimo i parametri sigma_j, j=1,...,J-1
199 sigma2.hat <- rep(NA,J)
200 for (j in 1:(J-1)) {
201   sigma2.hat[j] <- var(fP[,j], na.rm=T)
202 }
203 # utilizzo regressione log-lineare per stimare sigma_J
204 n <- length(sigma2.hat)
205 dev <- 1:n
206 my.dev <- dev[!is.na(sigma2.hat) & sigma2.hat > 0]
207 my.model <- lm(log(sigma2.hat[my.dev]) ~ my.dev)
208 sigma2.hat[is.na(sigma2.hat)] <- exp(predict(my.model, newdata=data
      .frame(my.dev=dev[is.na(sigma2.hat)])))
209
210 # stimo i parametri tau_j, j=1,...,J-2
211 tau2.hat <- rep(NA,J-1)
212 for (j in 1:(J-2)) {
213   tau2.hat[j] <- var(fI[,j], na.rm=T)
214 }
215 # utilizzo regressione log-lineare per stimare sigma_{J-1}
216 n <- length(tau2.hat)
217 dev <- 1:n
218 my.dev <- dev[!is.na(tau2.hat) & tau2.hat > 0]
219 my.model <- lm(log(tau2.hat[my.dev]) ~ my.dev)

```

```

220 tau2.hat[is.na(tau2.hat)] <- exp(predict(my.model, newdata=data.
      frame(my.dev=dev[is.na(tau2.hat)])))
221
222 # stimo parametri v2_j, j=1,...,J
223 v2 <- rep(NA,J)
224 for (i in 1:(J-1)) {
225   v2[i] <- sum(sigma2.hat) + sum(tau2.hat[i:J-1])
226 }
227 v2[J] <- sum(sigma2.hat)
228
229 # stimo parametri w2_j, j=1,...,J
230 w2 <- rep(NA,J)
231 for (i in 1:J) {
232   w2[i] <- sum(sigma2.hat[1:i])
233 }
234
235 # calcolo i parametri del vettore (c_1,...,c_J,b_1,...,b_{J-1})
236 c <- c()
237 for (j in 1:J) {
238   if (j==1) {
239     c[j] <- (1/sigma2.hat[j]) * sum(fP[1:J,j])
240   }
241   else if (j==2) {
242     c[j] <- (1/sigma2.hat[j]) * sum(fP[1:(J+1-j),j]) +
243       sum( 1/(v2[1:(j-1)]-w2[1:(j-1)]) *
244         log(tri2[J:(J-j+2),1:(j-1)]/tri[J:(J-j+2),1:(j-1)]))
245   } else {
246     diag2 <- diag(tri2[J:(J-j+2),1:(j-1)])
247     diag <- diag(tri[J:(J-j+2),1:(j-1)])
248     c[j] <- (1/sigma2.hat[j]) * sum(fP[1:(J+1-j),j]) +
249       sum( 1/(v2[1:(j-1)]-w2[1:(j-1)]) *
250         log(diag2[1:(j-1)]/diag[1:(j-1)]))
251   }
252 }
253
254 b <- c()
255 for (j in 1:(J-1)) {
256   if (j==1) {
257     b[j] <- -(1/tau2.hat[j]) * sum(fI[1:(J-j),j]) -
258       sum( 1/(v2[1:j]-w2[1:j]) *
259         log(tri2[J:(J-j+1),1:j]/tri[J:(J-j+1),1:j]))
260   } else {
261     diag2 <- diag(tri2[J:(J-j+1),1:j])
262     diag <- diag(tri[J:(J-j+1),1:j])
263     b[j] <- -(1/tau2.hat[j]) * sum(fI[1:(J-j),j]) -
264       sum( 1/(v2[1:j]-w2[1:j]) *
265         log(diag2[1:j]/diag[1:j]))
266   }
267 }

```

```

268
269 # calcolo l'inversa della matrice
270 # di varianza-covarianza a posteriori
271 a <- matrix(NA, nrow=(J + J - 1), ncol=(J + J - 1))
272 for (n in 0:(J-1)) {
273   for (m in 0:(J-1)) {
274     if(n==m) {
275       if (n==0) {
276         a[n+1,m+1] <- (J - n)/sigma2.hat[n+1]
277       } else {
278         a[n+1,m+1] <- ((J - 1) - n + 1)/sigma2.hat[n+1] +
279           sum(1/(v2[1:(min(n-1,m-1)+1)] -
280             w2[1:(min(n-1,m-1)+1)]))
281       }
282     } else {
283       if (n==0 | m==0) {
284         a[n+1,m+1] <- 0
285       } else {
286         a[n+1,m+1] <- sum(1/(v2[1:(min(n-1,m-1)+1)] -
287           w2[1:(min(n-1,m-1)+1)]))
288       }
289     }
290   }
291 }
292
293 for (n in 0:(J-2)) {
294   for (m in 0:(J-2)) {
295     if(n==m) {
296       a[J+n+1,J+m+1] <- (J-n-1)/tau2.hat[n+1] +
297         sum(1/(v2[1:min(n+1,m+1)] -
298           w2[1:min(n+1,m+1)]))
299     } else {
300       a[J+n+1,J+m+1] <- sum(1/(v2[1:min(n+1,m+1)] - w2[1:min(n+1,m+1)
301         ]))
302     }
303   }
304 }
305 for (n in 0:(J-1)) {
306   for (m in 0:(J-2)) {
307     if (n==0 | m==0) {
308       a[n+1,J+m+1] <- 0
309     } else {
310       a[n+1,J+m+1] <- -sum(1/(v2[1:(min(n-1,m)+1)] -
311         w2[1:(min(n-1,m)+1)]))
312     }
313   }
314 }
315

```

```

316 for (n in 0:(J-2)) {
317   for (m in 0:(J-1)) {
318     if (n==0 | m==0) {
319       a[J+n+1,m+1] <- 0
320     } else {
321       a[J+n+1,m+1] <- -sum(1/(v2[1:(min(n,m-1)+1)] -
322         w2[1:(min(n,m-1)+1)]))
323     }
324   }
325 }
326
327 # calcolo l'inversa della matrice calcolata
328 # ottenendo la matrice di
329 # varianza-covarianza a posteriori
330 ainvsolve(a)
331
332 # calcolo i parametri a posteriori
333 cb <- c(c,b)
334 theta.post <- ainvsolve %*% cb
335
336 # calcolo il vettore beta e s2.post che verranno
337 # utilizzati per il calcolo del valore atteso
338 # condizionato dei costi ultimi
339 beta <- c()
340 for (i in 1:(J-1)) {
341   beta[i] <- (v2[J] - w2[i])/(v2[i] - w2[i])
342 }
343
344 s2.post <- c()
345 E <- matrix(NA,nrow=(J-1),ncol=(2*J-1))
346 for (i in 2:J) {
347   e <- rep(0,J+1-i)
348   e <- c(e,rep(1 - beta[J+1-i],i-1))
349   e <- c(e,rep(0,J-i))
350   e <- c(e,rep(beta[J+1-i],i-1))
351   E[i-1,] <- e
352   s2.post[i-1] <- e %*% ainvsolve %*% e
353 }
354
355 # calcolo vettore dei costi ultimi
356 PIC.ult <- c()
357 for (i in 2:J) {
358   PIC.ult[i-1] <- tri[i,J+1-i]^(1 - beta[J+1-i]) *
359     tri2[i,J+1-i]^(beta[J+1-i]) * exp((1 - beta[J+1-i]) *
360     sum(theta.post[(J-i+2):J]) + beta[J+1-i] *
361     sum(theta.post[(2*J-i+1):(2*J-1)])) * exp((1 - beta[J+1-i]) *
362     (v2[J] - w2[J-i+1])/2 + s2.post[i-1]/2)
363 }
364

```

```

365 # calcolo riserve per anno di origine e riserva di portafoglio
366 PIC.Ris <- PIC.ult - diagP
367 PIC.RisTot <- sum(PIC.Ris)
368
369 # calcolo il MSEP
370 msep <- 0
371 for (i in 2:J) {
372   for (k in 2:J) {
373     if (i==k) {
374       msep <- msep + (exp((1-beta[J+1-i]) * (v2[J] - w2[J+1-i]) +
375         E[i-1, ]%% ainv %% E[k-1, ]) - 1) * PIC.ult[i-1] * PIC.ult[
376         k-1]
377     } else {
378       msep <- msep + (exp(E[i-1, ] %% ainv %% E[k-1, ]) - 1) *
379       PIC.ult[i-1] * PIC.ult[k-1]
380     }
381   }
382 }
PIC.se <- sqrt(msep)

```


Bibliografia

- [1] MACK, T. (1993). Distribution-free calculation of the standard error of chain ladder reserve estimates. *ASTIN Bulletin*, **23**(2), 213–225.
- [2] MERZ, M., WÜTHRICH, M. (2010). Paid–incurred chain claims reserving method. *Insurance: Mathematics and Economics*, **46**(3), 568–579.
- [3] PIGEON, M., ANTONIO, K. and DENUIT, M. (2013). Individual loss reserving with the multivariate skew normal framework. *ASTIN Bulletin*, **43**(3), 399–428.
- [4] PIGEON, M., ANTONIO, K. and DENUIT, M. (2014). Individual loss reserving using paid–incurred data. *Insurance: Mathematics and Economics*, **58**(C), 121–131.
- [5] QUARG, G., MACK, T. (2004). Munich chain ladder. *Blätter DGVM*, **XXVI**, 597–630.
- [6] WÜTHRICH, M., MERZ, M. (2008). *Stochastic Claims Reserving Methods in Insurance*, Wiley.